

494 TrpGluValLeuAlaArgHisGlySerLysIleTyrValAsnGluGlnPheLysLeuVal 513
1381 TGGGAGGTTTGGCAGGACCGCTCCAGATCTGGTCAATGAGGAGACCAAGCTGGTG 1440
514 TyrPheGlnGlyThrLysAspThrProLeuGluHisHisLeuTyrValValSerTyrGlu 533
1441 TACTTCCAGGACCAAGACACGCGCTGGAGCACCACTCTACGTGCTCAGCTATGAG 1500
534 AlaAlaGlyGluLeuValArgLeuThrThrProGlyPheSerHisSerCysSerMetSer 553
1501 GCGGCGCGGAGATCGCTACCGCTCCAGCGCGGTCTCCATAGCTGCTCCATGAGC 1560
554 GlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHis 573
1561 CAGAACTTCGACATGCTGCTAGCCACTACAGCAGCTGAGCAGCGCGCTGCGTGCAC 1620
574 ValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAla 593
1621 GTCTACAAGCTGAGCGCGCCGAGCAGACCCCTGCACAAGCAGCGCGCTCTGGGCT 1680
594 SerMetMetGluAlaAlaSerCysProProAspTyrValProGluIlePheHisPhe 613
1681 AGCATGATGAGGACCAACCGCTGCGCGGATATGTTCTCCAGAGATCTTCCATTTC 1740
614 HisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnPro 633
1741 CACACGCGCTCGATGCTGCGCTCTACGCGATGATCTACAAGCCCGCAGCTGCGAGCCA 1800
634 GlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsn 653
1801 GGGAGAAGCACCCACCGCTCTTGTATATGAGGCGCCCGCAGCTGCGTGAAT 1860
654 AsnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAla 673
1861 AACTCTTCAAGGCGATCAAGTACTTGGCGCTCAACACACTGCGCTCCCTGGGCTACGCC 1920
674 ValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeu 693
1921 GTGTTGTGATTCAGCGCAGGCGCTCTGTCAGCGAGGCTTCGGTTCGAGGGCGCCTG 1980
694 LysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAla 713
1981 AAAAACAATGGCCAGGTGGAGATCGAGGACCGAGGCGCGCTGCGTTCGTGGCC 2040
714 GluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrSerTyrGlyGly 733
2041 GAGAAGTATGGCTTCATCGACCTGAGCGAGTTGCCATCCATGCTGCTCTACGGGGC 2100
734 PheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGly 753
2101 TTCTCTCGCTATGGGCTTAATCCAAAGCCCGAGGTGTCAAGTGGCCATCGCGGGT 2160
754 AlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVal 773
2161 GCCCGGTACCGCTGATGGCTACGACACAGGGTACACTGAGCGCTACATGAGCGTC 2220
774 ProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeu 793
2221 CCTGAGAACAACCAAGCAGCGGTATGAGCGGGTTCCTGGCGCTGCGACGTGGAGAAGCTG 2280
794 ProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhe 813
2281 CCCAATGAGCCCAACCGCTGCTTATCTCCACGCGCTCTCCGAGCAAAAGCTGCATTT 2340
814 PheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGln 833
2341 TTCCACACAACTTCTCTCGTCTCCCACTGATCCGAGCAGGGAACCTTACCAGCTCCAG 2400
834 IleTyrProAsnGluArgHisSerIleArgCysProGluSerGlyGluHisTyrGluVal 853
2401 ATCTACCCCAACGAGACACAGATTCGCTGCCCGAGTCCGCGGCGAGCATTATGAAGTC 2460

854 ThrLeuLeuHisPheLeuGlnGluTyrLeu 863
2461 ACGTACTGCACTTCTACAGGATACCTC 2490
RESULT 8
AAI57896
ID AAI57896 standard; cDNA; 2801 BP.
XX AAI57896;
XX 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 99.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0489725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620317.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
P-PSDB; AAM38740.
XX Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
PS Claim 1; SEQ ID NO 99; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: immune system suppression,
activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemia and
C.N.S. diseases.
XX Sequence 2801 BP; 586 A; 889 C; 801 G; 525 T; 0 other;
Alignment Scores: 0 Length: 2801
Pred. No.: 0

Score: 4352.00 Matches: 827
Percent Similarity: 94.32% Conservative: 3
Best Local Similarity: 93.98% Mismatches: 5
Query Match: 93.67% Indels: 45
DB: 22 Gaps: 5

US-09-976-674-3 (1-863) x AAI57896 (1-2801)

QY 1 MetalThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspPro 20
~
Db 245 ATGGCCACACCGGACCCCAACAGCGCCGACGAGCGGACGCGCCACACATGACCG 304
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleHisGly 40
~
Db 305 GCCGCCGCTCCAGTGCAGAACACTCGTGGAGGGCTCCGGAGCATCATCCACGCG 364
QY 41 SerArgLysTrpSerGlyLeuIleValAsnLysAlaProHisAspPheGlnValGln 60
~
Db 365 AGCCCAAGTACTCGGCCCTCATTTGCAACAGGCGCCCGACGACTTCCAGTTTGGCAG 424
61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTrpLeuGlyMetProTrpGly 80
~
Db 425 AAGACGGATGACTGGGGCCCACTCCCAACCGCTCTACTACCTGGGTATGCATATGC 484
QY 81 SerArgGluAsnSerLeuLysSerGluIleProLysLysValArgLysGluAlaLeu 100
~
Db 485 ACCGAGAGAACTCCCTCTACTCTGAGATCCCAAGAGGTCGGAAGAGGCTCTG 544
QY 101 LeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisGlyVal 120
~
Db 545 CTGCTCTCTGCGAAGCAGATGCTGGATCATTTCCAGGCCACGCCCAACCTGGGTC 604
QY 121 TyrSerArgGluGluLeuLeuArgGluArgLysArgLysValPheGlyIleThr 140
~
Db 605 TACTCTCGGGAGGAGAGTGTGAGGGAGCGGAACGCTGGGGCTCTTCGGCATCAC 664
QY 141 SerTrpAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
~
Db 665 TCCTACGACTTCCACAGCAGAGTGGCTCTCTCTTCCAGGCCACGACGCTCTTC 724
QY 161 HisCysArgAspGlyLysAsnGlyPheMet-----ValSer 173
~
Db 725 CACTGCGCGAGCGGCGCAAGACGGCTTCATGGTGAGCGCTGGCCCTGGCTGTGTGC 784
QY 174 ProMetLysProLeuGluIleLysThrGlnCysSerGlyProArgMetAspProLysIle 193
~
Db 785 CCTATGAACCGCTGGAATCAAGACCCAGTGTCTAGGGCCCGGATGACCCCAAAATC 844
194 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsn 213
~
Db 845 TGCCCTGCGACCCCTGCTCTCTCTCTCATCAATAACAGCGACCTGTGGTGCGCAAC 904
QY 214 IleGluThrGlyGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 233
~
Db 905 ATCGACACAGCGGAGGAGCGGCGCTGACCTTCTGCCACCAAGGTTTATCAATGCTCG 964
QY 234 AspAspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPhe 253
~
Db 965 GATGACCCCAAGTCTCGGGGTGGGCCACTTCTGTATACAGGAAGTTCGACCGCTTC 1024
QY 254 ThrGlyTyrTrpTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeu 273
~
Db 1025 ACTGGGTACTGGTGGTGGCCACAGCCTCTCTGGAGAGGTTCAGAGGCCCTCAGACGCTG 1084
QY 274 ArgIleLeuTyrGluGluValAspGluSerGluValGlnValIleHisValProSerPro 293
~
Db 1085 CGAATCTGTATGAGGAAGTCGATGATCGAGGTGGAGTCAATTCACGTCCTCTCTCT 1144
QY 294 AlaLeuGluGluArgLysThrAspSerTrpArgTyrProArgThrGlySerLysAsnPro 313
~
Db 1145 GCCTAGAGAAGAAGAGCGGACTCGTATCGGTATCCCGCAGGACGACGACGAATCCC 1204
~
QY 314 LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 333

Db 1205 AAGATTGCTTGAACCTGGCTGAGTTCAGACTGACGCCAGGCGAGATCGTCTGACC 1264
QY 334 GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluTyrIleAla 353
~
Db 1265 CAGGAGAAGGAGCTGGTGCACCCCTTCAGCTCGCTGTTCCCGAGGTGAGTACATCCG 1324
QY 354 ArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGln 373
~
Db 1325 AGGCGCGGTGACCGCGGATGCAAAATACGCTGGGCATCTTCTCTGACCGGCCAG 1384
QY 374 GlnTrpLeuGlnLeuValLeuProProAlaLeuPheIleProSerThrGluAsnGlu 393
~
Db 1385 CAGTGGCTTCCAGCTCGTCTCTCCCGCGCCCTGTTCTATCCCGAGCAGAGATGAG 1444
QY 394 GluGln-ArgLeuAlaSerAlaArgAlaValProArgAsnValGlnProTrpTrpValVal 413
~
Db 1445 GAGCAGGCGGCTAGCTCTGCCAGAGCTGTCCCGAGGAATGTCCAGCGCTATGTGTGAC 1504
QY 413 rGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPheProGlnSe 433
~
Db 1505 GGAGGAGTCAACCAACGCTCTGGATCAATGTTCTATGACATCTTCTATCCCTTCCCAATC 1564
QY 433 rGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysH 453
~
Db 1565 AGAGGAGAGGAGGAGCTCTGCTTCTCCGCGCCCAATGAATGCAAGACCGGCTTCTGCA 1624
QY 453 sLeuTrpLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGluProPheSe 473
~
Db 1625 TTTGTACAAAGTCACCGCGCTTTAAATCCAGGGCTACGATTGGAGTGAGCCTTCAG 1684
QY 473 rProGlyGluAspGluPheLysCysProIleLysGluIleAlaLeuThrSerGlyG 493
~
Db 1685 CCCCAGGAAGGT-----GAGCAGACGCTGACGAATGCT-- 1718
QY 493 uTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluLysLeuVal 513
~
Db 1719 -----ATCTGGGTCAATGAGGAGACCAAGCTGGT 1747
QY 513 lTyrPheGlnGlyThrLysAspThrProLeuGluHisLysLeuTyrValValSerTrpG 533
~
Db 1748 GTACTTCCAGGCGACCAAGGACACCGCTGGAGCACCCTCTACGTGGTGCAGTATGA 1807
QY 533 uAlaAlaGlyIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSe 553
~
Db 1808 GCGGCGGCGAGATCGTACGCTCACCAGCCCGCTTCTCCCATAGTGTCTTCCATGAG 1867
QY 553 rGlnAsnPheAspMetPheValSerHisTyrSerValSerThrProProCysValH 573
~
Db 1868 CCAGAACTTCGACATGTCGTGACGCACTACAGAGCGTGAGCAGCCGCCCTCGGTGCA 1927
QY 573 sValTrpLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAl 593
~
Db 1928 CGTCTACAGCTGAGCGCGCCGAGCAGACCCCTGCAACAAGCAGCCCGCTTCTGGGC 1987
QY 593 aSerMetMetGluAlaAlaSerCysProProAspTrpValProProGluIlePheHisPh 613
~
Db 1988 TAGCATGATGGAGGAGCG-----AAGATCTTCATTT 2020
QY 613 eHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnPr 633
~
Db 2021 CCACAGCGCTCGGATGTGGGCTCTACGGCATGATCTACAAGCCCGCCGCTTCGAGCC 2080
QY 633 oGlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAs 653
~
Db 2081 AGGGAAGAAGCACCACCGCTCTCTTTGTATATGAGGCGCCCGCAGGTGCAGTGGTGAA 2140
QY 653 nasnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTrpAl 673
~
Db 2141 TAACCTCTTCAAGGCATCAAGTACTTGGGCTCAACACACTGGCCTCCCTGGGTACGC 2200
~
QY 673 aValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693

Db 2201 CGTGGTTGATTGACGGACGGGGCTCTGTACGACGAGGGCTTCGTTGGAAGGGCCCT 2260
 QY 693 uTysanaginMetGlyGlnValGluIleGluaspGlnValGluGlyLeuGlnPheValAl 713
 Db 2261 GAAAAACCAATGGCCAGATGGAGATCGAGACCAAGTGGAGGGCTGTGATTCGTGGC 2320
 QY 713 AGTGTGTYRGLYPheLeuAspLeuSerArgValAlaIleHisGlyTyrSerTyrGlyG 733
 Db 2321 CCGAGATATGGCTTCATGACCTGAGCCGACAGTTGCCATCCAGGGCTGGTCTTACGGGG 2380
 QY 733 yPheLeuSerLeuMetGlyLeuIleHisTyrSProGlnValPheLysValAlaIleAla 753
 Db 2381 CTTCCTCTGCTATGATGGGCTATTCACAAAGCCCGAGGTTCAGAGTGGCCATTCGGGG 2440
 QY 753 yAlaProValThrValTyrMetAlaTyrAspPheGlyTyrThrGluArgTyrMetAsp 773
 Db 2441 TGGCCCGGTACCGCTCGATGGCTTACGACAGGGGTACACAGGGCTACATGACGT 2500
 QY 773 lProGlnAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLys 793
 b 2501 CCTGAGAACACACAGCAGCTATGAGGGGTTCGTGGCCCTGACGCTGGAGAGCT 2560
 QY 793 uProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGlnAsnValHisP 813
 Db 2561 GCCCAATGACCCCAACCGCTTCTTATCTCCACGCTTCTGACCAAAAACGTCACCT 2620
 QY 813 ePheHisThrAspPheLeuValSerGlnLeuIleArgAlaGlyAspTyrGlnLeu-- 832
 Db 2621 TTTCACACAAATCTCTCGTCTCCAACTGATCCGAGCAGGAACCTTACACAGCTCCA 2680
 QY 833 -----GlnIleTyrProAsnGluArgHisSerIleArg 844
 Db 2681 GGTGGCCCTGCTCTGTCTCTCCCGCAGATTCACCCAAACAGAGACACATATTCGCTG 2740
 QY 844 sProGluSerGlyGlnHisTyrGluValThrLeuLeuHisPheLeuGlnGluTyrLeu 863
 Db 2741 CCCGAGTGGGGCAGCATTATGAGTACGTTGCTGACATTCTACAGAAATACCTC 2798
 RESULT 9
 ABRK3337
 ID ABRK3337 standard; cDNA; 4076 BP.
 AC ABRK3337;
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DPPP-2 splice variant #5.
 XX
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinnesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
 DR WPI: 2002-444178/47.
 DR P-PSDB; ABR61606.
 XX

PR New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PR the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PR viral infections, cancers, allergies, neurological disorders, or pain
 PS
 PS Disclosure: Page 90-91; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections,
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bullimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABRK3322-ABR83343 encode human DPPP proteins.
 CC
 SQ Sequence 4076 BP; 879 A; 1276 C; 1143 G; 778 T; 0 other;
 Alignment Scores:
 Pred. No.: 0 Length: 4076
 Score: 4337.00 Matches: 814
 Percent Similarity: 94.33% Conservative: 1
 Best Local Similarity: 94.21% Mismatches: 0
 Query Match: 93.35% Indels: 49
 DB: 24 Gaps: 1
 US-09-976-674-3 (1-863) x ABRK3337 (1-4076)
 QY 1 MetaLarThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 20
 Db 367 ATGGCCACACACCGGGAGCCCAAGCGCGAGCGCGAGCGCGACAGATGACCGG 426
 QY 21 AlaAlaArgPheGlnValGlnLysHisSerTyrAspGlyLeuArgSerIleIleHisGly 40
 Db 427 GCCCGCCGCTTCCAGGTGACAGAACCTCGTGGAGCGGCGCCGAGCATATCCACGGC 486
 QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
 Db 487 AGCCGACAGTACTGGGCTCATTTGTCACAAAGCGCCCGACGATTCATGTTGTGAG 546
 QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
 Db 547 AAGACGATGAGTCTGGGCCCCACCTCCACCGCGCTACTACGGAATGCCATATGCG 606
 QY 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
 Db 607 AGCCGAGAGAACTCCCTCTACTCTGATGATTCACAAAGAGTCCGGAAGAGCTCTG 666
 QY 101 LeuLeuLeuSerTyrPheGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
 Db 667 CTGCTCTGCTCTGGAACAGATGCTGATTCATTCAGGCCACGCCCAACATGGGGCT 726
 QY 121 TyrSerArgGluGluGluLeuLeuArgLysValArgLysValPheGlyIleThr 140
 Db 727 TACTCTGGGAGGAGGAGCTGCTGAGGAGGGAACCGCTGGGGTCTTGGCATCAC 786
 QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
 Db 787 TCCTACGACTTCCACAGGAGAGTGGCTTCTTCCAGGCGCAACAGCCCTCTTC 846
 QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
 Db 847 CACTCCCGGAGCGGCGGCAAGACGCTTCATGAGTGTCCCTCATGAACCGCTGAAATC 906
 QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
 Db 907 AAGACCAAGTGTCTAGGGGCCCGGAGTGAACCCCAAAATCTGCCCTCCAGACCTGCTTC 966


```

|||||
Db 653 TyGlyGlyProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArg 672
QY 2000 CTCACACACTGGCTCCCTGGGCTACGCCGCTGTGTATGATGAGCGAGGCTCTGT 2059
Db 673 LeuAsnThrLeuAlaSerLeuGlyTyrValValValValIleAspAsnArgGlySerGys 692
QY 2060 CACGAGAGGCTGGTTCGAGAGGCGCCGTAACCAATGGCCAGGTGAGATCGAG 2119
Db 693 HisArgGlyLeuLysPheGlnGlyAlaPheLysTyrLysMetGlnIleGlnLeuAsp 712
QY 2120 GACCAGGTGAGGCGCTGACAGTCTGTGCGCGAGAGTATGCTCATGACCTGAGCCGA 2179
Db 713 AspGlnValGlnGlyLeuGlnTyrLeuAlaSerArgTyrAspPheIleAspLeuAspArg 732
QY 2180 GTTGCCATCATGGCTGTGCTTACGGGCGCTCTCTGCTCATGAGGCTATCCACAG 2239
Db 733 ValGlyLeuHisGlyTyrPheSerTyrGlyGlyTyrLeuSerLeuMetAlaLeuMetGlnArg 752
QY 2240 CCCCAGGTGTTCAGGTTCAGGCTCATCGCGGCTGCGCGTACCGCTGAGATGGCTACGAC 2299
Db 753 SerAspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAsp 772
QY 2300 ACAGGCTACACTGAGCGCTACATGAGAGCTCCCTGAGAACACACAGCGCTATGAGCG 2359
Db 773 ThrGlyTyrThrGlnArgTyrMetGlnHisProAspGlnAsnGlnGlnIleTyrTyrLeu 792
QY 2360 GGTTCGCTGGCGCTGACGAGTGGAGAGCTGCCAATGAGCCCAACCGCTGCTTCTCTC 2419
Db 793 GlySerValAlaMetGlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeu 812
QY 2420 CACGCTTCCTGACGAGAAAGCTGACCTTTTTCACACAACTTCCTGCTCTCCACATG 2479
Db 813 HisGlyPheLeuAspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
QY 2480 ATCCGAGCAGAGGAAACCTTACACAGCTCCAGATCTACCCCAACAGACACAGATTTGCC 2539
Db 833 ValArgAlaGlyLysProGlyTyrAspLeuGlnIleTyrProGlnArgHisSerIleArg 852
QY 2540 TGCCCGAGTGGCGGAGCAGCTATGAGTACAGTGTGCTGCTCATGAGGATACCTC 2599
Db 853 ValProGlnSerGlyGlnHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnLysLeu 872

RESULT 4
075273 PRELIMINARY: PRT; 508 AA.
AC 075273:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE R26984.1 (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SOURCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Atlix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coeffield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Komiyoshi A., Arellano A., Montgomey M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carraro A.V.,
RT "Sequence analysis of a 2.5 kb region in 19p13.3."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005594; AAC33801.1;
DR MEROPS; S09.019;
DR InterPro; IPR002410; PRO_annotase.
DR InterPro; IPR000379; Ser_estrs_site.
DR PRINTS; PRO0793; PROAMNPTASE.
NON_TER 1 1

```

```

SQ. SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;
Alignment Scores:
Pred. No.: 3,36e-163 Length: 508
Score: 2572.00 Matches: 493
Percent Similarity: 91.34% Conservative: 3
Best Local Similarity: 90.79% Mismatches: 3
Query Match: 53.64% Indels: 44
DB: 4 Gaps: 5

US-09-976-674-4 (1-2617) x 075273 (1-508)
QY 998 ATGCTGTGACCCAGAGAGAGAGAGAGTGTGTGACCCCTTACAGTCCGCTTCCGAGAGTG 1057
Db 1 IleValSerThrGlnGlnGlnLysGlnLeuValGlnProPheSerLeuPheProLysVal 20
QY 1058 GAGTACATGCCAGGCGCGGCTGACCCGCGGATGCAATATACGCTGGGCGCATTTCTG 1117
Db 21 GlnTyrIleAlaArgAlaGly-----AlaThrPalaMetPheLeu 33
QY 1118 GACCGGCGCCAGACAGTGGCTCAGCTGCTCTCTCCCGCGCTTTCATCCGAGAC 1177
Db 34 AspArgProGlnGlnTyrPheGlnLeuValLeuLeuProAlaLeuPheIleProSer 53
QY 1178 ACAGAGATGAGAGACAGCAGGCTAGCCTGTGCGAGCTGTCCCGAGAGATGTCACCGG 1237
Db 54 ThrGlnAsnGlnGlnGlnArgLeuAlaSerAlaAlaValProAlaAsnValGlnPro 73
QY 1238 TATGTGTGTACAGAGAGTACCAAGCTGTGATCAATGTTCATGACATTTCTATCC 1297
Db 74 TyrValValTyrGlnGlnValThrAsnValThrIleAsnValHisAspIlePheTyrPro 93
QY 1298 TTCCCGCATGAGAGAGAGAGAGAGAGTGTGCTTCCCGCGCATGAAATGCAAGAC 1357
Db 94 PheProGlnSerGlnGlnGlnLysPheLeuLysPheLeuArgAlaAsnGlnLysTyr 113
QY 1358 GCGTCTGTGCAATTTGTACAAAGTACCGCGCTTTAAATCCAGGCTACGATTTGAGAT 1417
Db 114 GlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTyrSer 133
QY 1418 GAGCGCTTACCGCGCGGAGAGATGAAATTAAGTCCCGCATGAGAGATGCTGTG 1477
Db 134 GluProPheSerProGlnGlnGly-----GlnGlnSerLeu 145
QY 1478 ACCAGCGGTGAATGGAGAGTTTGGCGAGCGAGCTCAAGATCTGGTCAATGAGAG 1537
Db 146 ThrAsnAla-----IleThrPalaAsnGlnGln 154
QY 1538 ACCAGCTGTGTACTTCCAGGCGACCAAGACACCGCGCTGAGACACCACTCTACGTG 1597
Db 155 ThrLysLeuValTyrPheGlnGlnTyrLysAspThrProLeuGlnHisLeuTyrVal 174
QY 1598 GTACAGTATGAGCGCGCGGAGATGCTGACGCTCACACGCGCGCTTCCATACG 1657
Db 175 ValSerTyrGlnAlaAlaGlyGlnIleValArgLeuThrThrProGlyPheSerHisSer 194
QY 1658 TGCCTCATGAGCGCAAACTGTGACATGTGCTGACCGCACTACAGACGTGACACCGCG 1717
Db 195 CysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSerValSerThrPro 214
QY 1718 CCTCGCTGACAGCTTAAAGTGAAGCGCGCGAGACGACGCGCGCGCAAGCGAGCGCC 1777
Db 215 ProGlyValHisValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnPro 234
QY 1778 CGCTTCTGGGCTAGCATATGAGAGACAGCTGCCCGCGGATTAATGTCTCTCCAGAG 1837
Db 235 ArgPheThrPalaSerMetMetGlnAlaAla-----Lys 245
QY 1838 ATCTTCATTTCCACACCGCGCTCGAGTGTGCGCTTACGAGCATGATTCAGAGCGCCAC 1897
Db 246 IlePheHisPheHisThrArgSerAspValArgLeuTyrGlnMetIleTyrLysProHis 265
QY 1898 GCGTTGACAGCCAGGAGAGAGACACCCACCGTCTTTGTATATGAGAGCGCCAGAGTG 1957

```



```

|||||
Db 266 AlaleuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyLysProGlnVal 285
1958 CAGGTGGTGAATATCTCTCCCAAGGACATCACTCTGGGCTCAACACTGGCTCC 2017
Db 286 GlnLeuValAsnAsnSerPheLysGlyLeuTyrLeuArgLeuAsnThrLeuAlaSer 305
2018 CTGGGCTACGGCCGTGTGTGATTGACGAGGAGGCTCTCTGACGAGGCTTGGTTC 2077
Db 306 LeuGlyTyrAlaValValValLeuAspGlyArgGlySerCysLysIleArgGlyLeuArgPhe 325
2078 GAAGGGCCCTGAAAAACCAATGGCCAGGTGAGATCGAGACAGGACAGGTGAGGGCTG 2137
Db 326 GluGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnAspGlnValGlnGlyLeu 345
2138 CAGTTCGTGGCCGAGATGATGAGCTTCATGACCTGAGCTGAGCTGACCTGAGCTGG 2197
Db 346 GlnPheValAlaGlnLysTyrGlyPheLeuAspLeuSerIleArgAlaAlaLeuHisGlyTyrP 365
2198 TCGTACGGGGGCTCTCTGCTGCTCATGGGCTAATCCACAAGCCCCAGGTGTCAAGGTG 2257
Db 366 SerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysVal 385
2258 GCGATCGGGGGTGGCCGGGTACCGCTGATGAGCTGACGACAGGAGTACCTGAGCGC 2317
Db 386 AlaIleAlaGlyAlaProValThrValThrPheAlaTyrAspThrGlyTyrThrGlnArg 405
2318 TACATGACGCTCCCTGAGAACACACAGCAGCAGGCTATGAGGCGGGTCCGTGGCCCTGAC 2377
Db 406 TyrMetAspValProGlnAsnAsnGlnHisGlyTyrGlnAlaGlySerValAlaLeuHis 425
2378 GTGGAGAAAGCTCCCAATGAGACCCCAACCGCTGTATCTCCACAGGCTTCCGAGACGA 2437
Db 426 ValGlyLysLeuProAsnGlnProAsnArgLeuLeuIleHisGlyPheLeuAspGln 445
2438 AACGCTACTTTTCCACACAACCTCTGCTCCCAACTGATCCGAGAGGGAACCT 2497
Db 446 AsnValHisPhePheHisThrAsnPheLeuValSerIleLeuArgAlaGlyLysPro 465
2498 TACACAGCTC-----CAGATCTACCCCAACAGAGACAC 2530
Db 466 TyrGlnLeuGlnValAlaLeuProProValSerProGlnIleTyrProAsnGlnAlaGlnHis 485
2531 AGTATTCGCTGCGCCGAGTGGGAGAGACACTAATGACTCACTGTCGCTTACAG 2590
Db 486 SerIleArgCysProGlnSerCylGlnHisTyrGlnValThrLeuLeuHisPheLeuGln 505
2591 GAATACGCTC 2599
Db 506 GluTyrLeu 508

```

RESULT 5

```

09BVR3 PRELIMINARY; PRT; 439 AA.
AC 09BVR3:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical 49.9 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000970; AA00970.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR00379; Ser_estra_site.
DR Pfam; PF00326; Peptidase_S9; 1.

```

```

KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;
Alignment Scores:
Pred. No.: 1,28e-150 Length: 439
Score: 2383.00 Matches: 439
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49.70% Indels: 0
DB: 4 Gaps: 0
US-09-976-674-4 (1-2617) x 09BVR3 (1-439)
QY 1283 GACATCTTATCCCTTCCCAATGAGGAGAGAGAGAGAGTCTGCTTCCGCGCC 1342
Db 1 AspIlePheThrProPheProGlnSerGlnGlyLysArgGlnLeuCysPheLeuAla 20
QY 1343 AATGATGCAAGACCGGCTTCCCATTTGTACAAAGTCACCGCGTTTAAATCCAG 1402
Db 21 AsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGln 40
QY 1403 GGCTACGATGAGATGAGCCCTTCAAGCCCGGGGAAAGTGAATTGAATGATGATTAAG 1462
Db 41 GlyTyrAspThrSerIleThrProPheSerProGlyGlnAspGlnPheLysCysProIleLys 60
QY 1463 GAAGAGATTGCTGTGACACAGCGGTGATGGAGGATTTGGCGAGGACGCTCCAGATC 1522
Db 61 GluGlnIleAlaLeuThrSerGlyGlnTyrGlnValLeuAlaArgHisGlySerLysIle 80
QY 1523 TGGGTCAATGAGGAGACCAAGCGGTGTACTCCAGGGGACCAAGACCGCTGGAG 1582
Db 81 TrpAlaSerGlnGlnThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuGln 100
QY 1583 CACACCTCTACCTGTGTCAAGTATGAGGCGGCGGAGATGTACGCTCAGACAGGCC 1642
Db 101 HisHisLeuTyrValValSerTyrGlnAlaAlaGlyLysIleValArgLeuThrThrPro 120
QY 1643 GCGTCTCCCATAGCTGTCCATAGAGCCAGAACTCGACATGTGTGACGACATACAGC 1702
Db 121 GlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSer 140
QY 1703 AGCTGAGCAGCGCGCGCTGCTGACGTACAGCTGACAGCTGAGCGCCCGAGACAGCCC 1762
Db 141 SerValSerThrProGlnCysValHisValTyrLysLeuSerGlyProAspArgPro 160
QY 1763 CTGACACAAGACCGCGCTTCTGGCTACGATGATGAGAGCGACGCTGCGCCGAGAT 1822
Db 161 LeuHisLysGlnProArgPheThrAlaSerMetGlnAlaAlaSerCysProProAsp 180
QY 1823 TATGTTCTCCAGAGATCTCCATTTCCACACGCGCTGAGTGTGCGCTACAGCATG 1882
Db 181 TyrValProProGlnIlePhePheHisThrArgSerAspValArgLeuTyrIleMet 200
QY 1883 ATCTACAGGCCCAAGCGCTTGTGACGAGGAGACACCCACCGCTCTTTGTATAT 1942
Db 201 IleTyrLysProHisAlaLeuGlnProGlyLysHisProThrValLeuPheValIle 220
QY 1943 GGAGGCCCCAGGAGGAGCTGGTGAATTAAGCTCTTAAGGATCAAGTACTGGGGCTC 2002
Db 221 GlyLysProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArgLeu 240
QY 2003 AACACACTGGCGCTCCCTGGGCTACGCGCTGTGTGATTGACGAGGAGGCTCTGTAG 2062
Db 241 AsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCysGln 260
QY 2063 CGAGGCTGTGGCTGCAAGGAGGCGCTGAAAAACCAATGGCCAGGTGAGATGAGAGNC 2122
Db 261 ArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnLys 280
QY 2123 CAGGTGAGGCGCTGTGATTCGTGGCGAGAAAGTATGCTTATGACCTGAGCGGAGTT 2182
Db 281 GlnValGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAspLeuSerArgVal 300

```


Db 312 AspaspProLysSerAlaGlyValAlaThrPheValIleGlnGlnPheAspArgPhe 331
 QY 770 ACTGGTACTGTGGTGGCCACAGCCCTCGGGAAGGTTCAAGAGGCCCTCAAGAGCTG 829
 Db 332 ThrGlyTrpTrpPheCysProThrAlaSerTrpGlu-----GlnGlyLeuYsthrLeu 349
 QY 830 CGAATCCTGTATGAGGAAGTGGATGAGTCCAGAGTGGAGTCAATGACGCCCTCTCT 889
 Db 350 ArgIleLeuTrpGlnGlnValAlaSpGlnSerGlnValGlnValIleHisValProSerPro 369
 QY 890 GCGCTAGAGAAAGAGAGAGAGAGTCTGATGCGTACCCAGGACAGCAAGAAATCC 949
 Db 370 AlaLeuGlnGlnArgLysTrpAspSerTrpArgTrpProArgTrpGlySerLysAsnPro 389
 QY 950 AAGATTGCTTGAAGTGGCTGAGTCCAGACTGACAGCCAGGCAAGATTCCTGACC 1009
 Db 390 LysIleAlaLeuLysLeuAlaGlnPheGlnThrAspSerGlnGlyLysIleValSerThr 409
 QY 1010 CAGAGAGAGAGAGTGGTGGCAGCCCTTACGCTGCTGCCAAGGTGAGATGATCGCC 1069
 Db 410 GlnGlnLysGlnLeuValGlnProPheSerSerLeuPheProLysValGlnTrpIleAla 429
 QY 1070 AGGGCCGGG 1078
 Db 430 ArgAlaGly 432
 RESULT 8
 Q96NT8 PRELIMINARY; PRT; 312 AA.
 ID Q96NT8 AC Q96NT8;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE CDNA FLJ30094 f1s, clone BNGH4100034, weakly similar to dipeptidyl
 DE peptidase IV (EC 3.4.14.5).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
 RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kikuchi K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
 RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
 RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.,
 RT "MEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK054656; BAB70784.1;
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR00379; Ser esters_site.
 DR Pfam: PF00326; Peptidase_S9; 1
 SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC5877717 CRC64;

QY 1784 TGGGCTAGCATGfTGGAGGAGGACAGCTGCCCGGATATGTTCCACAGATCTTC 1843
 Db 41 TrpAlaSerMetMetGlnAlaAlaSerCysProPheAspTrpValProGlnIlePhe 60
 QY 1844 CATTTCCAGCGGCGTGGATGGCGGCTTACGGCAGATGATTAACACCCAGGCTTC 1903
 Db 61 HisPheHisThrArgSerAspValAlaGlyLeuTrpGlyMetIleTrpLysProHisAlaLeu 80
 QY 1904 CAGCCAGGGAAGAGACCCACCCGCTCTTGTATATGAGAGGCCCCAGGTGACAGT 1963
 Db 81 GlnProGlyLysLysHisProThrValLeuPheValTrpGlyGlyProGlnValGlnLeu 100
 QY 1964 GTGATTAACCTTCAAGGCAATCAAGTACTTGGCGGCTCAACACACTGCTCCCTGGGC 2023
 Db 101 ValAsnAsnSerPheLysGlnLysTrpLeuAlaGlyLeuAsnThrLeuAlaSerLeuGly 120
 QY 2024 TACCGCGTGGTGTGATGATGACGAGGCGCTCTGTCAGCGAGGCTCGGTTGGAAGG 2083
 Db 121 TyrAlaValAlaValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGlnGly 140
 QY 2084 GCGCTGAAAACCAATGGGCCAGGTGGAGATCGAGAACAGGTGAGGCGCTCCAGTTC 2143
 Db 141 AlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnAspGlnValGlnGlyLeuGlnPhe 160
 QY 2144 GTGGCCGAGAGATGATGCTTCAATCGACCTGAGCGAGTGGCATGCGATGGCTGCTAC 2203
 Db 161 ValAlaGlnLysTrpGlyLysPheLeuAspSerAlaValAlaIleHisGlyTrpSerTrp 180
 QY 2204 GGGGCGTCTCTGCTCATGGGCGCTATCGAACAGCCCGAGGTGTCAGGTGGCCATC 2263
 Db 181 GlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 200
 QY 2264 GCGGAGCGCCCGCTGACCTGATGGGCTTACGAGAGGTACAGGCTACAGGCGCTATAGT 2323
 Db 201 AlaCysAlaProValAlaThrValTrpMetAlaTrpAspTrpGlyTrpGlnArgTrpMet 220
 QY 2324 GACGTCCTGAGAACACAGCAGCGCTATGAGGCGGCTCGAGCCCTGACAGTGGAG 2383
 Db 221 AspValProGlnAsnAsnGlnHisGlyTrpGlnAlaGlySerValAlaLeuHisValGln 240
 QY 2384 AACGTCGCCAATGAGGCCAAGCGCTTGTATCCGCCAGGCTCTCCGAGAAAGATG 2443
 Db 241 LysLeuProAsnGlnProAsnArgLeuLeuIleHisGlyPheLeuAspGlnAsnVal 260
 QY 2444 CACTTTTCCACAAACTTCTGCTCCCAACTGATCCAGCAGGAAACTTACAG 2503
 Db 261 HisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGln 280
 QY 2504 CTCACATCTACCCCAAGCAGAGACAGATATGCTGCTCCCGGAGTGGCGAGCACTAT 2563
 Db 281 LeuGlnIleTrpProAsnGlnTrpGlnHisSerIleArgCysProGlnSerGlyGlnHisTrp 300
 QY 2564 GAAGTCACGTTGCTGCTCACTTCTACAGAAATACCTC 2599
 Db 301 GlnValThrLeuLeuHisPheLeuGlnGlnTrpLeu 312
 RESULT 9
 Q9VC20 PRELIMINARY; PRT; 1042 AA.
 ID Q9VC20 AC Q9VC20;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE CG3744 protein.
 GN CG3744.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 11:53:37 ; Search time 34 Seconds
(without alignments)
3382.214 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: .4646
Sequence: 1 MATTGTPRADRGDAATDPD.....CPESGHEVETLLHFLQEV 863

Scoring table:
BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4646	100.0	863	23	ABG61592 Human DPPIV relate
2	4646	100.0	892	23	ABG61602 Human DPP-2 splic
3	4646	100.0	892	23	ABG61604 Human DPP-2 splic
4	4636	99.8	899	23	AAE24168 Human dipeptidyl p
5	4558.5	98.1	879	23	ABG61607 Human DPP-2 splic
6	4558.5	98.1	879	23	ABG61608 Human DPP-2 splic
7	4458	96.0	830	23	AAE24171 Human dipeptidyl p
8	4279	92.1	869	23	AAE24169 Alternative versio
9	4129	88.9	847	23	AAE23875 Murine dipeptidyl
10	4026	86.7	832	23	ABG61605 Human DPP-2 splic

11	4026	86.7	832	23	ABG61606	Human DPP-2 splic
12	3938.5	84.8	819	23	ABG61609	Human DPP-2 splic
13	3938.5	84.8	819	23	ABG61610	Human DPP-2 splic
14	3817.5	82.2	737	22	AAE24172	Human polypeptide
15	3539.5	76.2	683	22	AAE24173	Human polypeptide
16	3409.5	73.4	720	21	AAE24174	Human OREF ORF1390
17	2870	61.8	882	22	AAE24175	Human DPP-2 splic
18	2870	61.8	882	22	AAE24176	Human DPP-2 splic
19	2870	61.8	882	22	AAE24177	Human DPP-2 splic
20	2870	61.8	882	22	AAE24178	Human DPP-2 splic
21	2870	61.8	882	22	AAE24179	Human DPP-2 splic
22	2688	57.9	497	23	ABG61615	Human DPP-2 splic
23	2688	57.9	497	23	ABG61616	Human DPP-2 splic
24	2684	57.8	497	23	ABG61617	Human DPP-2 splic
25	2684	57.8	497	23	ABG61618	Human DPP-2 splic
26	2572	55.4	508	21	AAE24180	Human secreted pro
27	2562	55.1	518	21	AAE24181	Human secreted pro
28	2547	54.8	518	21	AAE24182	Human secreted pro
29	2414	52.0	782	23	ABG61603	Human DPP-2 splic
30	2178	46.9	724	23	ABG61604	Human DPP-2 splic
31	2137.5	46.0	632	22	AAE24183	Human DPP-2 splic
32	2109.5	45.4	460	22	AAE24184	Human DPP-2 splic
33	2103.5	45.3	512	22	AAE24185	Human DPP-2 splic
34	1998	43.0	690	23	ABG61594	Human DPP-2 splic
35	1997	43.0	658	23	ABG61595	Human DPP-2 splic
36	1997	43.0	661	23	ABG61596	Human DPP-2 splic
37	1816	39.1	613	23	ABG61601	Human DPP-2 splic
38	1628.5	35.1	1042	22	AAE24186	Human DPP-2 splic
39	1628.5	35.1	1102	22	AAE24187	Human DPP-2 splic
40	1583	34.1	580	23	AAE24188	Human DPP-2 splic
41	1494.5	32.2	465	22	AAE24189	Human DPP-2 splic
42	1220.5	26.3	360	22	AAE24190	Human DPP-2 splic
43	1093.5	23.5	310	22	AAE24191	Human DPP-2 splic
44	1093.5	23.5	310	22	AAE24192	Human DPP-2 splic
45	1047	22.5	349	23	ABG61602	Human DPP-2 splic

ALIGNMENTS

RESULT 1	ABG61592	standard; Protein: 863 AA.
ID	ABG61592	
XX	12-AUG-2002 (first entry)	
AC	ABG61592	
XX	12-AUG-2002 (first entry)	
DT	12-AUG-2002 (first entry)	
XX	Human DPPIV related serine protease DPP-2.	
XX	Human: serine protease: dipeptidyl peptidase IV-related protein: DPP-2.	
KW	DPPIV: Infection; human immunodeficiency virus; HIV-1; HIV-2; pain;	
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;	
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;	
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;	
KW	metabolic disorder; reproductive disorder; inflammatory disorder;	
XX		
OS	Homo sapiens.	
PN	WO20023134-A2.	
XX		
PD	18-APR-2002.	
XX		
PF	12-OCT-2001; 2001WO-US31874.	
XX		
PR	12-OCT-2000; 2000US-240117P.	
XX		
PA	(FERR) FERRING BV.	
XX		
PI	Ol S, Akinsanya KO, Riviere PJ, Junien J;	
XX	WPI; 2002-44178/A7.	

DR N-PSDB; ABR83323.
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 PS Claim 17, Fig 1: 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPP-IV proteins.
 CC
 XX
 XX Sequence 863 AA:
 S0
 Query Match 100.0%; Score 4646; DB 23; Length 863;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATTGPTADGDAADDDPAARFQVQKHSMDGLRSIIHGRKYSGLIVNKAHPDFQVQ 60
 DB 1 MATTGPTADGDAADDDPAARFQVQKHSMDGLRSIIHGRKYSGLIVNKAHPDFQVQ 60
 QY 61 KTDSSGSHSLYLLGMPYGRSLSLYSELPKVRKALLLSMKOMLDFQATPHHGV 120
 DB 61 KTDSSGSHSLYLLGMPYGRSLSLYSELPKVRKALLLSMKOMLDFQATPHHGV 120
 QY 121 YSREELLREKRLGVGITSYDFHSESGFLFQASNSLFCRQDGGKGFVSPMKPLEI 180
 DB 121 YSREELLREKRLGVGITSYDFHSESGFLFQASNSLFCRQDGGKGFVSPMKPLEI 180
 QY 121 YSREELLREKRLGVGITSYDFHSESGFLFQASNSLFCRQDGGKGFVSPMKPLEI 180
 DB 181 KTCQSGPRMDKICPADPAFFSFITNSDLWVANIETGEERLTCFHOGLSVLDDPKSAG 240
 QY 241 VAFVIOEEDRFPTGYMWCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 DB 241 VAFVIOEEDRFPTGYMWCPTASWEGSGKTLRLIYEVDSEVEVLIHVPSPALEEKT 300
 QY 301 DSYRPTGSKNPKIALKLAFFQDSOGKIYSTOKELVQFSSLPKVEYIARAGWTRD 360
 DB 301 DSYRPTGSKNPKIALKLAFFQDSOGKIYSTOKELVQFSSLPKVEYIARAGWTRD 360
 QY 361 GKIVAMFLDRPQOMLOLVLLPALFTPTSENEBORLASARAVPRNOPYVVEEVNTNW 420
 DB 361 GKIVAMFLDRPQOMLOLVLLPALFTPTSENEBORLASARAVPRNOPYVVEEVNTNW 420
 QY 421 INVNDITFPFQSGEGELCFILRANECKTGFCCHLYKVAVALKSGCYDMSSEFSGEDEK 480
 DB 421 INVNDITFPFQSGEGELCFILRANECKTGFCCHLYKVAVALKSGCYDMSSEFSGEDEK 480
 QY 481 CPIKEETALTGEMEVILARHSGKIWNNEETKLYVFOGTQKDPLEHLHYVVEAAGELVR 540
 DB 481 CPIKEETALTGEMEVILARHSGKIWNNEETKLYVFOGTQKDPLEHLHYVVEAAGELVR 540
 QY 541 LTPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRFWASMEERAS 600
 DB 541 LTPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRFWASMEERAS 600
 QY 601 CRRPYVPEETIFHFTSRDVRILYGMILYKPHALQPKKHPTVLFVGGQVOLVNNSEFKIK 660
 DB 601 CRRPYVPEETIFHFTSRDVRILYGMILYKPHALQPKKHPTVLFVGGQVOLVNNSEFKIK 660

QY 661 YLRNLTLASIGYAVVVIDGRSCORGLRFEGALKNGMGVEIEDVQGLQFPAEKYGFID 720
 DB 661 YLRNLTLASIGYAVVVIDGRSCORGLRFEGALKNGMGVEIEDVQGLQFPAEKYGFID 720
 QY 721 LSRVAIHGWSYGGFLSLMGLHKPQVKVATAGAPVTYVMAAYDTGTYERTYVDEENNQHG 780
 DB 721 LSRVAIHGWSYGGFLSLMGLHKPQVKVATAGAPVTYVMAAYDTGTYERTYVDEENNQHG 780
 QY 781 YEAGSVLHVKEKLPNEPRRLILHGFIDENYHFTNPLVSQILRACKPYOLOIYPERH 840
 DB 781 YEAGSVLHVKEKLPNEPRRLILHGFIDENYHFTNPLVSQILRACKPYOLOIYPERH 840
 QY 841 SIRCPESGEHYEVTLLHFLQEYL 863
 DB 841 SIRCPESGEHYEVTLLHFLQEYL 863
 RESULT 2
 ABG61602
 ID ABG61602 standard; Protein: 892 AA.
 XX
 AC ABG61602;
 XX
 DT 12-AUG-2002 (first entry)
 DE
 XX Human DPP-2 splice variant #1.
 XX
 KW Human: serine protease; dipeptidyl peptidase IV-related protein; DPP-IV;
 KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 KW
 XX Homo sapiens.
 XX
 OS WO200231134-A2.
 PN
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Juntun J;
 XX
 DR WPI: 2002-444178/47.
 DR
 XX N-PSDB; ABR83333.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 XX
 PS Disclosure; Page 76-78; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
 CC proteins (DPP-IV). The dipeptidyl peptidase IV-related proteins (DPP-IV)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and

CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	
XX	
Sequence	892 AA;
Query Match	100.0%; Score 4646; DB 23; Length 892;
Best Local Similarity	100.0%; Pred. No. 0;
Matches	863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MATGTPTADRDGAATDDPAARFOVOKHSMGDLRSIIHSGRSKYSGLLVNKAHPHFOFOVQ 60
DB	30 MATGTPTADRDGAATDDPAARFOVOKHSMGDLRSIIHSGRSKYSGLLVNKAHPHFOFOVQ 89
QY	61 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKREKALLLSMKOMLDHFOATPHHGV 120
DB	90 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKREKALLLSMKOMLDHFOATPHHGV 149
QY	121 YSREBELLRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFMVSPMKPLEI 180
DB	150 YSREBELLRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFMVSPMKPLEI 209
QY	181 KTQCSGPRMDPKICPADPAFFSFINNSDLVANIETGEERRLTFCOGLSNVLDPPKSG 240
DB	210 KTQCSGPRMDPKICPADPAFFSFINNSDLVANIETGEERRLTFCOGLSNVLDPPKSG 269
QY	241 VATEVIOREPFRTGYMWCPTASMEGSEGLKTLRLILEVDESEVEVLIHVPSPALPEERT 300
DB	270 VATEVIOREPFRTGYMWCPTASMEGSEGLKTLRLILEVDESEVEVLIHVPSPALPEERT 329
QY	301 DSYRPTGTGKNPKIALKLAEFQTDGKIVSTOEKELVQPFSSLPKVEYIARAGWTRD 360
DB	330 DSYRPTGTGKNPKIALKLAEFQTDGKIVSTOEKELVQPFSSLPKVEYIARAGWTRD 389
QY	361 GKVAMAMLDPRQOMLOLVLLPPLAFITSTENEBQRLASARAVRNNQPYVVEVTVW 420
DB	390 GKVAMAMLDPRQOMLOLVLLPPLAFITSTENEBQRLASARAVRNNQPYVVEVTVW 449
QY	421 INVADIEFPFQSEGEDELCLFRANECKTGFCCHLYKVTAVLKSQGYDMSEPPSPEDERK 480
DB	450 INVADIEFPFQSEGEDELCLFRANECKTGFCCHLYKVTAVLKSQGYDMSEPPSPEDERK 509
QY	481 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFQGTDPRLHNHLYVSYEAGELIVR 540
DB	510 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFQGTDPRLHNHLYVSYEAGELIVR 569
QY	541 LTTFGFSHSCMSQNFDMFVSHSVSTPCVHYKILSGPDDPLHKQPRFASMEAS 600
DB	570 LTTFGFSHSCMSQNFDMFVSHSVSTPCVHYKILSGPDDPLHKQPRFASMEAS 629
QY	601 CPPDVPEPEIFHFTSRDVRILYGMILYKPHALOPGKKHPTVLFVYGGPOVOLVNNSEFKGIK 660
DB	630 CPPDVPEPEIFHFTSRDVRILYGMILYKPHALOPGKKHPTVLFVYGGPOVOLVNNSEFKGIK 689
QY	661 YLRNLTLASLIGYAVVVVDGRSGCQGLRFEGLKNOMQOVETEDVEGLQFAEERYGFLD 720
DB	690 YLRNLTLASLIGYAVVVVDGRSGCQGLRFEGLKNOMQOVETEDVEGLQFAEERYGFLD 749
QY	721 LSRVAIHMSVGGFLSLMGLIHKROYEKVATAGAPVYVMAVDTYTERKYDVPENNCHG 780
DB	750 LSRVAIHMSVGGFLSLMGLIHKROYEKVATAGAPVYVMAVDTYTERKYDVPENNCHG 809
QY	781 YEAGSVALHVKLEPNEPRRLILHGFLEPENHFFHTNLFVLSOLIRAGRPYOLQIYPERNH 840
DB	810 YEAGSVALHVKLEPNEPRRLILHGFLEPENHFFHTNLFVLSOLIRAGRPYOLQIYPERNH 869
QY	841 SIRPESGEHYEVTLLHFLQEYL 863
DB	870 SIRPESGEHYEVTLLHFLQEYL 892

RESULT 3
ABG61604
ID ABG61604 standard; Protein: 892 AA.

AC	ABG61604;
XX	
XX	
DT	12-NOV-2002 (first entry)
DE	
XX	
XX	
KW	Human DPRP-2 splice variant #3.
KW	DPRP; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW	DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW	dyskinesia; reproductive disorder; inflammatory disorder;
KW	metabolic disorder.
OS	
XX	Homo sapiens.
XX	
PN	WO200231134-A2.
PD	18-APR-2002.
XX	
XX	
PE	12-OCT-2001; 2001WO-US31874.
PR	12-OCT-2000; 2000US-240117P.
XX	
PA	(FERR) FERRING BV.
PI	Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX	
DR	WPI; 2002-444178/47.
XX	
XX	N-PSDB; ABK83335.
PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT	viral infections, cancers, allergies, neurological disorders, or pain
XX	
PS	Disclosure: Page 81-84; 113pp; English.
XX	
CC	The present invention relates to the isolation of novel human serine
CC	proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC	proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC	and nucleic acids encoding them are useful for treating infections
CC	such as fungal, bacterial, protozoan and viral infections, particularly
CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC	bulimia, Parkinson's disease, acute heart failure, hypertension,
CC	hypertension, urinary retention, osteoporosis, angina pectoris,
CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC	psychotic and neurological disorders (e.g. anxiety, dementia, or
CC	schizophrenia), and dyskinesias. These may also be used in discovering
CC	therapeutic agents for the treatment of reproductive, inflammatory and
CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	
XX	
Sequence	892 AA;
QY	
DB	
DB	
QY	1 MATGTPTADRDGAATDDPAARFOVOKHSMGDLRSIIHSGRSKYSGLLVNKAHPHFOFOVQ 60
DB	30 MATGTPTADRDGAATDDPAARFOVOKHSMGDLRSIIHSGRSKYSGLLVNKAHPHFOFOVQ 89
QY	61 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKREKALLLSMKOMLDHFOATPHHGV 120
DB	90 KTDESGPHSHRLYLGMYPGSRNSLXSEIPKVKREKALLLSMKOMLDHFOATPHHGV 149
QY	121 YSREBELLRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFMVSPMKPLEI 180
DB	150 YSREBELLRERKRLGVFGITSDYDFHSESGFLFOASNSLFHCRDGGKNGFMVSPMKPLEI 209
QY	181 KTQCSGPRMDPKICPADPAFFSFINNSDLVANIETGEERRLTFCOGLSNVLDPPKSG 240

```

Db 210 KTQCSGPRMDPKICPADPAFFSFITNSDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 269
QY 241 VATEVIOEEFDRFTGYWMCPTASMEGSEGLKTLRILEYEVDSEVEVYIHVSPALAEKRT 300
Db 270 VATEVIOEEFDRFTGYWMCPTASMEGSEGLKTLRILEYEVDSEVEVYIHVSPALAEKRT 329
QY 301 DSYRYPRGSKNPKIALKLAIEFQTDGSKIVSTQEKELVQFSSLPKVEYIARAGWTRD 360
Db 330 DSYRYPRGSKNPKIALKLAIEFQTDGSKIVSTQEKELVQFSSLPKVEYIARAGWTRD 389
QY 361 GKYAMAMFLDRPQOMLOLVLLPALFTIPSTENEBORLASARAVPRNVPYVVEEVNTNW 420
Db 390 GKYAMAMFLDRPQOMLOLVLLPALFTIPSTENEBORLASARAVPRNVPYVVEEVNTNW 449
QY 421 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEFSPGDEEFK 480
Db 450 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEFSPGDEEFK 509
QY 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGTRDPLEHHLTVVSYEAAGEIVR 540
Db 510 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGTRDPLEHHLTVVSYEAAGEIVR 569
QY 541 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRPMASMEAS 600
Db 570 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRPMASMEAS 629
QY 601 CPPDYVPELTFHFTSRDVRLYGMIYKPHALQPKKHPTVLFVYGGPOVOLVNNSEFKIK 660
Db 630 CPPDYVPELTFHFTSRDVRLYGMIYKPHALQPKKHPTVLFVYGGPOVOLVNNSEFKIK 689
QY 661 YLRINTLASLQYAAVVYIDGRSCQGRLEFGALKNMQGOVEIEDQVEGLQFAEKYGFID 720
Db 690 YLRINTLASLQYAAVVYIDGRSCQGRLEFGALKNMQGOVEIEDQVEGLQFAEKYGFID 749
QY 721 LSRVAIHQMSVGGFLSLMGLLHKPOVEKVALAGAPVYVMAVYDGYTRYVDVPENNHG 780
Db 750 LSRVAIHQMSVGGFLSLMGLLHKPOVEKVALAGAPVYVMAVYDGYTRYVDVPENNHG 809
QY 781 YEAGSVLHVKEKLPNEPRRLILHGFLENVHFHTNPLVLSQLIRAGKPYOLOIYPRNRH 840
Db 810 YEAGSVLHVKEKLPNEPRRLILHGFLENVHFHTNPLVLSQLIRAGKPYOLOIYPRNRH 869
QY 841 SIRCESEGEHYEVLHFLQEYL 863
Db 870 SIRCESEGEHYEVLHFLQEYL 892

RESULT 4
AAE24168
ID AAE24168 standard; Protein: 969 AA.
XX
AC AAE24168;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 9 (DPP9) protein.
XX
KM Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KM autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KM graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KM antiviral; enzyme.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT Misc-difference 374
XX /note= "Encoded by GAA"
XX WO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.

```

```

XX 27-OCT-2000; 2000AU-0001078.
PR (UNSY ) UNIV SYDNEY.
PA
XX
PI Abbott CA, Gorrell MD;
XX
DR WPI; 2002-454646/48.
DR N-PSDB; AAD38954.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Claim 1; Fig 4; 91pp; English.
XX
CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP9 protein.
XX
SQ Sequence 969 AA;
XX
Query Match 99.8%; Score 4636; DB 23; Length 969;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MATTGTPADRGDAATDDPARFOYOKHSMWGLRSIIHGSRYKSLIVNKAPHDQFQVQ 60
Db 107 MATTGTPADRGDAATDDPARFOYOKHSMWGLRSIIHGSRYKSLIVNKAPHDQFQVQ 166
QY 61 KTDSGSPSHSLYLYLGMFYGSRNSLSYSELPKVKYRKALLLSMKQMLDHFQATPRHGV 120
Db 167 KTDSGSPSHSLYLYLGMFYGSRNSLSYSELPKVKYRKALLLSMKQMLDHFQATPRHGV 226
QY 121 YSRREELLREKRLGVGRTSYDSHSESGFLPFOASNSLFCRGRGKNGFVSPMKPLEI 180
Db 227 YSRREELLREKRLGVGRTSYDSHSESGFLPFOASNSLFCRGRGKNGFVSPMKPLEI 286
QY 181 KTQCSGPRMDPKICPADPAFFSFITNSDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 240
Db 287 KTQCSGPRMDPKICPADPAFFSFITNSDLWVANIETGEERLTFCHQGLSNVLDDPKSAG 346
QY 241 VATEVIOEEFDRFTGYWMCPTASMEGSEGLKTLRILEYEVDSEVEVYIHVSPALAEKRT 300
Db 347 VATEVIOEEFDRFTGYWMCPTASMEGSEGLKTLRILEYEVDSEVEVYIHVSPALAEKRT 406
QY 301 DSYRYPRGSKNPKIALKLAIEFQTDGSKIVSTQEKELVQFSSLPKVEYIARAGWTRD 360
Db 407 DSYRYPRGSKNPKIALKLAIEFQTDGSKIVSTQEKELVQFSSLPKVEYIARAGWTRD 466
QY 361 GKYAMAMFLDRPQOMLOLVLLPALFTIPSTENEBORLASARAVPRNVPYVVEEVNTNW 420
Db 467 GKYAMAMFLDRPQOMLOLVLLPALFTIPSTENEBORLASARAVPRNVPYVVEEVNTNW 526
QY 421 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEFSPGDEEFK 480
Db 527 INVHDIFFPPOSGEDELCLFLRANECKTGCFCHLYKTVAVLKSOGYDMSSEFSPGDEEFK 586
QY 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGTRDPLEHHLTVVSYEAAGEIVR 540
Db 587 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGTRDPLEHHLTVVSYEAAGEIVR 646
QY 541 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRPMASMEAS 600
Db 647 LITPGFSHSCMSQNFDMFVSHYSSVSTPCVHYKLSGPDDDLHKOPRPMASMEAS 706
QY 601 CPPDYVPELTFHFTSRDVRLYGMIYKPHALQPKKHPTVLFVYGGPOVOLVNNSEFKIK 660
Db 707 CPPDYVPELTFHFTSRDVRLYGMIYKPHALQPKKHPTVLFVYGGPOVOLVNNSEFKIK 766

```


Page 5

Qy	661	YLNLTLTASLGVAVVVDIGRSGCQGLRFEGLAKNQMGVEIEDVEGLQIAEAYGTFID	720
Db	767	YLRLLTLTSLSLGAVVAVDIDGRSGCGLRFEGLAKNQMGVEIEDVEGLQIAEAYGTFID	826
Qy	721	LSRVALHGSVSGEFLSLMGLIHKPOVFPVVALAGAVPYWMAVDQGYTERMYDPENNOHG	780
Db	827	LSRVALHGSVSGEFLSLMGLIHKPOVFPVVALAGAVPYWMAVDQGYTERMYDPENNOHG	886
Qy	781	YEAGSVALHVEKLPEPNRNLTLHGFEDENVAFFFTNLFVSOLILNAGRPYQLOIIPNERSH	840
Db	887	YEAGSVALHVEKLPEPNRNLTLHGFEDENVAFFFTNLFVSOLILNAGRPYQLOIIPNERSH	946
Qy	841	SIRCPESGSHYEVYTLHLHQLQETL	863
Db	947	SIRCPESGSHYEVYTLHLHQLQETL	969
RESULT 5			
ABG61607	ID	ABG61607 standard; Protein; 879 AA.	
XX	AC	ABG61607;	
XX	DT	12-AUG-2002 (first entry)	
XX	DE	Human DPPR-2 splice variant #6.	
XX	KW	Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;	
KW	KW	DPPV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;	
KW	KW	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;	
KW	KW	heart failure; hypertension; urinary retention; osteoporosis; cancer;	
KW	KW	ulcer; allergy; cancer; psychotic disorder; neurological disorder;	
KW	KW	dyskinesia; reproductive disorder; inflammatory disorder;	
XX	OS	metabolic disorder.	
XX	OS	Homo sapiens.	
XX	PN	WO200231134-A2.	
XX	PD	18-APR-2002.	
XX	PF	12-OCT-2001; 2001WO-US31874.	
XX	PR	12-OCT-2000; 2000US-240117P.	
XX	PA	(FERR) FERRING BV.	
XX	PI	Qi S, Akinsanya KO, Riviere PJ, Junien J;	
XX	XX	WPI: 2002-444178/47.	
XX	DR	N-PSDB; ABR83338.	
XX	PT	New dipeptidyl peptidase IV-related proteins and nucleic acids encoding	
XX	PT	the proteins, useful for treating e.g. fungal, bacterial, protozoan and	
XX	PT	viral infections, cancers, allergies, neurological disorders, or pain	
XX	PS	-	
XX	PS	Disclosure; Page 91-93; 113pp; English.	
XX	CC	The present invention relates to the isolation of novel human serine	
XX	CC	proteases referred to as dipeptidyl peptidase IV (DPPV)-related	
XX	CC	proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)	
XX	CC	and nucleic acids encoding them are useful for treating infections	
XX	CC	such as fungal, bacterial, protozoan and viral infections, particularly	
XX	CC	infections caused by human immunodeficiency virus (HIV-1 or HIV-2),	
XX	CC	pain, diabetes, precocious puberty, infertility, obesity, anorexia,	
XX	CC	bulimia, Parkinson's disease, acute heart failure, hypotension,	
XX	CC	hypertension, urinary retention, osteoporosis, angina pectoris,	
XX	CC	stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,	
XX	CC	psychotic and neurological disorders (e.g. anxiety, dementia, or	
XX	CC	schizophrenia), and dyskinesias. These may also be used in discovering	
XX	CC	therapeutic agents for the treatment of reproductive, inflammatory and	

CC	metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX	
Sequence	879 AA;
Query Match:	98.1%; Score 4558.5; DB 23; Length 879;
Best Local Similarity	98.5%; Pred. No. 0;
Matches 850; Conservative	0; Mismatches 0; Indels 13; Gaps 1;
1	MATGTPTADRDAAATDPPARFQVOKHSMDSGLSIHSGSKYSGLIVNKAAPHDFQVQ 60
30	MATGTPTADRDAAATDPPARFQVOKHSMDSGLSIHSGSKYSGLIVNKAAPHDFQVQ 89
61	KTDESGPSSHRLYLIGMPYSGRENSLSEIRPKVKREALLLSWKQMLDHFQATPHNGV 120
90	KTDESGPSSHRLYLIGMPYSGRENSLSEIRPKVKREALLLSWKQMLDHFQATPHNGV 149
121	YSREELLRRRRRLGVFGITSYDFHSESGFLFQASNSLFFCRQGGKNGFVSPKPLEI 180
150	YSREELLRRRRRLGVFGITSYDFHSESGFLFQASNSLFFCRQGGKNGFVSPKPLEI 209
181	KTCGSGPRMDPKICPADPAFFSFINNSDLVMAINITEGERRLTFCHQSLSNVLDPRKAG 240
210	KTCGSGPRMDPKICPADPAFFSFINNSDLVMAINITEGERRLTFCHQSLSNVLDPRKAG 269
241	VATFVIOEEDRFTGYWMCPTASWEGSGELKTLRIEYDEVESEVEYLIVHPSPALEERT 300
270	VATFVIOEEDRFTGYWMCPTASWEGSGELKTLRIEYDEVESEVEYLIVHPSPALEERT 329
301	DSYRTPRPGSKPKITALKLAERQTSQGLVSTQEKELVORPSSLFPRVEYIARAGWTRD 360
330	DSYRTPRPGSKPKITALKLAERQTSQGLVSTQEKELVORPSSLFPRVEYIARAGWTRD 389
361	GKYAAMFLDRQOQLVLLPLPALEFISTENEERILASARVLRPNVORVYVEEVTVW 420
390	GKYAAMFLDRQOQLVLLPLPALEFISTENEERILASARVLRPNVORVYVEEVTVW 449
421	INWHDIEYFPQSEDEDELCPFLANECKTGCHLKVTAVLKSQGYDMSPEPSEDEPK 480
450	INWHDIEYFPQSEDEDELCPFLANECKTGCHLKVTAVLKSQGYDMSPEPSEDEPK 509
481	CPKEERIALTSGEWEVLARHGSKIWNEETKLIVFOGTRKPRLENNHLLVVSYEAGELIV 540
510	CPKEERIALTSGEWEVLARHGSKIWNEETKLIVFOGTRKPRLENNHLLVVSYEAGELIV 556
541	LTPPGSHSCSOMNDFMVSHYSSVSPRCVHYKLSGDDDDLYHKOPREFASMEEAS 600
557	LTPPGSHSCSOMNDFMVSHYSSVSPRCVHYKLSGDDDDLYHKOPREFASMEEAS 616
601	CPDDVYPRPIFFPHTRSDVRLYGMITYKPHALOPGKKHPVULFVYGGPOVLVNNFSFKIK 660
617	CPDDVYPRPIFFPHTRSDVRLYGMITYKPHALOPGKKHPVULFVYGGPOVLVNNFSFKIK 676
661	YLRLNTLASLGAUVVLDGRGSCQGLRFEGLAKNOMQVELEDQVEGLQFAEKYGEID 720
677	YLRLNTLASLGAUVVLDGRGSCQGLRFEGLAKNOMQVELEDQVEGLQFAEKYGEID 736
721	LSRVALHGSYGGFSLIMGLIHKPOVFKALAGAYTYMAADDTGYTERKYMVPERNNHG 780
737	LSRVALHGSYGGFSLIMGLIHKPOVFKALAGAYTYMAADDTGYTERKYMVPERNNHG 796
781	YEAGSVALVEKLPRNEPNLLILHGFIDENVAFFPTNFLVSOLI RAGRPYOLQIYPRNERH 840
797	YEAGSVALVEKLPRNEPNLLILHGFIDENVAFFPTNFLVSOLI RAGRPYOLQIYPRNERH 856
841	SIRCPESGEHYEVTLLHFLQEYL 863
857	SIRCPESGEHYEVTLLHFLQEYL 879

```

AC ABG61608;
XX
XX 12-AUG-2002 (first entry)
XX
XX Human DPRP-2 splice variant #7.
DE
XX
XX Human: serine protease: dipeptidyl peptidase IV-related protein: DPRP;
KW DPPIV; Infection: human immunodeficiency virus: HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinestia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.
XX
XX Homo sapiens.
OS
XX WO200231134-A2.
PN
XX 18-APR-2002.
PD
XX 12-OCT-2001; 2001WO-US31874.
PF
XX
XX 12-OCT-2000; 2000US-240117P.
PR
XX
XX (FERR ) FERRING BV.
PA
XX
XX OI S, AKinsanya KO, Riviere PJ, Junien J;
PI
XX WPI: 2002-444178/47.
DR
XX N-PSDB; ABK83339.
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
XX Disclosure; Page 95-96; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
XX
SQ Sequence 879 AA;

```

```

Query Match 98.1%; Score 4558.5; DB 23; Length 879;
Best local Similarity 98.5%; Pred. No. 0;
Matches 850; Conservative 0; Mismatches 0; Indels 13; Gaps 1;

```

```

Db 210 KTQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGERRLTFCHOGLSNVLDPPKSAG 269
QY VATEVIOEEDPFRFTGYMWCPTASWEGSGELKTLRLIYEVEDESEVEVHVSPALEEKT 300
Db 270 VATEVIOEEDPFRFTGYMWCPTASWEGSGELKTLRLIYEVEDESEVEVHVSPALEEKT 329
QY DSYRVPRTGSKNPKIATLAEFGQDSQKIVSTQEKELVQPFSSLPKVEYIARAGWTRD 360
Db 330 DSYRVPRTGSKNPKIATLAEFGQDSQKIVSTQEKELVQPFSSLPKVEYIARAGWTRD 389
QY 361 GKYVAMAMLDPRQOMLOLVLLPPLAFIPSTENEGORLASAAVARNQPVYVVEVYVW 420
Db 390 GKYVAMAMLDPRQOMLOLVLLPPLAFIPSTENEGORLASAAVARNQPVYVVEVYVW 449
QY 421 INVHDIFFPQSGEDDELCTFLRANECKTGCHLYKYTAVALKSOGYDMSEPFSGEDEFK 480
Db 450 INVHDIFFPQSGEDDELCTFLRANECKTGCHLYKYTAVALKSOGYDMSEPFSGEDEFK 509
QY 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVYFQGTDPLEHHLVYSYEAGELIVR 540
Db 510 CPIKEEIALTSGEWEVLARHGSK-----GTRKDPLEHHLVYSYEAGELIVR 556
QY 541 LTTGCFSHSCSMSONFDKFFVSHYSVSTRPCVHYVYKLSGPDDDLHKQPRFMAEMEAAS 600
Db 557 LTTGCFSHSCSMSONFDKFFVSHYSVSTRPCVHYVYKLSGPDDDLHKQPRFMAEMEAAS 616
QY 601 CPPDYVPEEIFPHFTRSDVRLYGMITYKPHALQPKKHPTVLFVYGGPOVOLVNNSPFKIK 660
Db 617 CPPDYVPEEIFPHFTRSDVRLYGMITYKPHALQPKKHPTVLFVYGGPOVOLVNNSPFKIK 676
QY 661 YLRNLTASLIGYAAVVVIDGRSCQGLRFEKALKNOMQVEIEDQVEGLQFAEKYGFID 720
Db 677 YLRNLTASLIGYAAVVVIDGRSCQGLRFEKALKNOMQVEIEDQVEGLQFAEKYGFID 736
QY 721 LSRVAIHGMSYSGFLSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERKYVDVEENQHG 780
Db 737 LSRVAIHGMSYSGFLSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERKYVDVEENQHG 796
QY 781 YEAGSVALHVKLNEPRLILHGFIDENVHFPTNPLVSQLRAGKPYOLOIYPNERH 840
Db 797 YEAGSVALHVKLNEPRLILHGFIDENVHFPTNPLVSQLRAGKPYOLOIYPNERH 856
QY 841 SIRCESEGEHYEVTLHFLQEYL 863
Db 857 SIRCESEGEHYEVTLHFLQEYL 879

```

RESULT 7
AAE24171
ID AAE24171 standard; Protein: 830 AA.
XX
AC AAE24171;
XX
DT 23-SEP-2002 (first entry)
XX
XX Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.
DE
XX
KW Human: dipeptidyl peptidase: DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW graft rejection; antidiabetic; antiinflammatory; immunosuppressive;
KW antiviral; enzyme; DPP-4 like 2 protein.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 235 /note= "Encoded by GAG"
FT
XX
XX WO200234900-A1.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.
PF

• Fri Dec 13 12:15:05 2002

```

XX 27-OCT-2000: 2000AU-0001078.
PR
XX
XX (UNSY ) UNIV SYDNEY.
XX
XX Abbott CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX N-PSDB: AAD38957.
XX
XX New dipeptidyl peptidase (DPP) peptidases, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX PS Disclosure; Page 82-86; 91pp; English.
XX
XX CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX CC polynucleotides encoding such proteins. The DPP peptidases are useful for
XX CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX CC rejection and HIV (human immuno deficiency virus) infection. The present
XX CC sequence is human DPP4-like 2 protein.
XX
XX SQ Sequence 830 AA;
XX
XX Query Match 96.0%; Score 4458; DB 23; Length 830;
XX Best Local Similarity 99.8%; Pred. No. 0;
XX Matches 828; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 34 LRSTIHSGSRKYSGLIVNKAHPDFOVOKTDESGPHSHRLYYLGMFYSGRENSLTYSEIPK 93
XX 1 LRSTIHSGSRKYSGLIVNKAHPDFOVOKTDESGPHSHRLYYLGMFYSGRENSLTYSEIPK 60
XX
XX QY 94 KVRKEALLLSWKOMDLHFOATPHHGVYSREELLREKRRLGVGITSYDHSSEGLFLF 153
XX 61 KVRKEALLLSWKOMDLHFOATPHHGVYSREELLREKRRLGVGITSYDHSSEGLFLF 120
XX
XX QY 154 QASNSLPHCRDGGKNGFMVSPMKLEIKTQCSGPRMDKTCBPADAPAFSFINNSDLWYAN 213
XX 121 QASNSLPHCRDGGKNGFMVSPMKLEIKTQCSGPRMDKTCBPADAPAFSFINNSDLWYAN 180
XX
XX QY 214 IETGEERRLTFCHQGLSNVLDPPKSAGVATFVIOEEDRFYGMWCPYASWEGSEGLTKL 273
XX 181 IETGEERRLTFCHQGLSNVLDPPKSAGVATFVIOEEDRFYGMWCPYASWEGSEGLTKL 240
XX
XX QY 274 RILVEEVDESEVEVIHVPSPALERTDSTRYPRTGSKNPKIALKLAFOQDSOGKIVST 333
XX 241 RILVEEVDESEVEVIHVPSPALERTDSTRYPRTGSKNPKIALKLAFOQDSOGKIVST 300
XX
XX QY 334 QEKELVOPFSSLFPKVEYIAAGWTRDGGKVMAMFLDRPOQWLDLVLLPPLFTPSTENE 393
XX 301 QEKELVOPFSSLFPKVEYIAAGWTRDGGKVMAMFLDRPOQWLDLVLLPPLFTPSTENE 360
XX
XX QY 394 EORLASAAVPRNQQPVYVEEVNWNVINWADIFYPPQSEGEDELCTFLANECKTGCH 453
XX 361 EORLASAAVPRNQQPVYVEEVNWNVINWADIFYPPQSEGEDELCTFLANECKTGCH 420
XX
XX QY 454 LYKTAVALKSGQIDWSEPFSGDEDFKCPKIEEIALTSGEWEVLARHGSKIWNVEETKLV 513
XX 421 LYKTAVALKSGQIDWSEPFSGDEDFKCPKIEEIALTSGEWEVLARHGSKIWNVEETKLV 480
XX
XX QY 514 YFOGTKTPTLEHNLXYVSYEAGEIVRLTPGFSHSCSMQNFPMFVSHSVSTPQVH 573
XX 481 YFOGTKTPTLEHNLXYVSYEAGEIVRLTPGFSHSCSMQNFPMFVSHSVSTPQVH 540
XX
XX QY 574 VYKLSGDDPLHAKOPRFWASMMMEASCPDYVPPEITFHFTTRSDVRLGYGITYRPHALQ 633
XX 541 VYKLSGDDPLHAKOPRFWASMMMEASCPDYVPPEITFHFTTRSDVRLGYGITYRPHALQ 600
XX
XX QY 634 GKRRPTVLFYVGGPOVOLVNNSEFKGYTLRLNTLASLGAVAVVIDGRSGCORGLRFEGL 693
XX 601 GKRRPTVLFYVGGPOVOLVNNSEFKGYTLRLNTLASLGAVAVVIDGRSGCORGLRFEGL 660

```

```

QY 694 KNMGQVEIEDQVEGLQFAEKYGFIDLSRYAIGHMSYGGFLSLMGLTHKPOVEKVALAT 753
DB 661 KNMGQVEIEDQVEGLQFAEKYGFIDLSRYAIGHMSYGGFLSLMGLTHKPOVEKVALAT 720
QY 754 APVTVMAYDYGYTERYKDVDPENNOHGYEAGSVALLHVEKLNEPNRRLIILGFLDENYHF 813
DB 721 APVTVMAYDYGYTERYKDVDPENNOHGYEAGSVALLHVEKLNEPNRRLIILGFLDENYHF 780
QY 814 FHTNFIQSILRAGKPYOLQIYPNDRHSIRCPESGEHEVTLHFLOEYL 863
DB 781 FHTNFIQSILRAGKPYOLQIYPNDRHSIRCPESGEHEVTLHFLOEYL 830

RESULT 8
AAE24169
ID AAE24169 standard; Protein; 869 AA.
XX
XX AC AAE24169;
XX
XX DT 23-SEP-2002 (first entry)
XX
XX DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) protein.
XX
XX KM Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
XX KM human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
XX KM type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
XX KM antiviral; enzyme.
XX
XX OS Mus sp.
XX
XX PN WO200234900-A1.
XX
XX PD 02-MAY-2002.
XX
XX PF 29-OCT-2001; 2001WO-AU01388.
XX
XX PR 27-OCT-2000; 2000AU-0001078.
XX
XX PA (UNSY ) UNIV SYDNEY.
XX
XX PI Abbott CA, Gorrell MD;
XX
XX PS Claim 1; Page 70-74; 91pp; English.
XX
XX CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX CC polynucleotides encoding such proteins. The DPP peptidases are useful for
XX CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX CC rejection and HIV (human immuno deficiency virus) infection. The present
XX CC sequence is an alternative version of murine DPP9 protein.
XX CC Note: This sequence is stated to be the same as that shown as
XX CC SEQ ID NO: 4 in the sequence listing of the specification. However these
XX CC sequences differ.
XX
XX SQ Sequence 869 AA;
XX
XX Query Match 92.1%; Score 4279; DB 23; Length 869;
XX Best Local Similarity 92.1%; Pred. No. 0;
XX Matches 792; Conservative 32; Mismatches 36; Indels 0; Gaps 0;
XX
XX QY 4 TGTPTADRCDAATDDPARFOVQKHSMDGLRSTIHSGSRKYSGLIVNKAHPDFOVOKTD 63
XX 10 SGVSPVEQVAAGMDMDTARFCVOVKHSMDGLRSTIHSGSRKSSGLIVSKAHPDFOVOKPD 69

```

```

QY 64 ESGPHSHRLYYLGMYPGSRNSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 123
DB 70 ESGPHSHRLYYLGMYPGSRNSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 129
QY 124 EEBLLRRKRRLGVGITSYDFHSGSLFLFOANSFLHCRDGGKNGFMVSPMKPLEIKTQ 183
DB 130 EEBLLRRKRRLGVGITSYDFHSGSLFLFOANSFLHCRDGGKNGFMVSPMKPLEIKTQ 189
QY 184 CSGPRMDPKICPADPAFFSFINSNDLVANIENTGEERLFLFCHOGLSNVLDPPKSAGVAT 243
DB 190 CSGPRMDPKICPADPAFFSFINSNDLVANIENTGEERLFLFCHOGLSNVLDPPKSAGVAT 249
QY 244 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVYIHVSPALBERKTDY 303
DB 250 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVYIHVSPALBERKTDY 309
QY 304 RYPTGSKNPKIALKLAELFOTDSOGKITVTOEKELVOPFSSLPFKVYIARAGTTRGKY 363
DB 310 RYPTGSKNPKIALKLAELFOTDSOGKITVTOEKELVOPFSSLPFKVYIARAGTTRGKY 369
QY 364 AMAMFLDRPOOMLOLVLLPPALFIPSTENEBOQLASARAARVPRNQPYYVEEVTNWINV 423
DB 370 AMAMFLDRPOOMLOLVLLPPALFIPSTENEBOQLASARAARVPRNQPYYVEEVTNWINV 429
QY 424 HDIFPRPOSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSQGYDWSPEFGEDEFKCP 483
DB 430 HDIFPRPOSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSQGYDWSPEFGEDEFKCP 489
QY 484 KEELALTSGEWEVLARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVRLTT 543
DB 490 KEELALTSGEWEVLARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVRLTT 549
QY 544 PGFSHSCSMSONDFMFVSHSVSTPCVHYKLSGPDPLHFKOPRFMASMMEASCP 603
DB 550 LGFSHSCSMSONDFMFVSHSVSTPCVHYKLSGPDPLHFKOPRFMASMMEASCP 609
QY 604 DYPPPELFHHTKSDVLYLGMITKRPALDQCKKHPTVLYGGRQVOLVNNSEFGITLYR 663
DB 610 DYPPPELFHHTKSDVLYLGMITKRPALDQCKKHPTVLYGGRQVOLVNNSEFGITLYR 669
QY 664 LNTLASIGYAVVYIDGSGSCORGREFGALKNGOYEDQVEGLDFVAKYGFIDLSR 723
DB 670 LNTLASIGYAVVYIDGSGSCORGREFGALKNGOYEDQVEGLDFVAKYGFIDLSR 729
QY 724 VALHGSYGGFLSLMGLIHKRQVFKVAIAGAPVTVMAYDTGYTERYMDYPENNOGYEA 783
DB 730 VALHGSYGGFLSLMGLIHKRQVFKVAIAGAPVTVMAYDTGYTERYMDYPENNOGYEA 789
QY 784 GSAVALHEKLPNEPNRLLILHGFLEENVHFFHTNFVLSOLIRAGKPYOLOIYPERHSIR 843
DB 790 GSAVALHEKLPNEPNRLLILHGFLEENVHFFHTNFVLSOLIRAGKPYOLOIYPERHSIR 849
QY 844 CPESGEHEVTLHLHLEOYL 863
DB 850 CPESGEHEVTLHLHLEOYL 869

```

```

XX XX WO200234900-A1.
PN
XX
XX 02-MAY-2002.
PD
XX
XX 29-OCT-2001; 2001WO-AU01388.
PF
XX
XX 27-OCT-2000; 2000AU-0001078.
PR
XX
XX (UNSY ) UNIV SYDNEY.
PA
XX
XX Abbott CA, Gorrell MD;
PI
XX
XX WPI; 2002-45464/48.
DR
XX
XX N-PSDB: AAD38311.
XX
PT New dipeptidyl peptidase (DPP) peptides, useful for screening
PT Inhibitors of DPP catalytic activity, which may be employed to treat
PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
PT rejection and HIV infection -
XX
PS Claim 1; Fig 8; 91pp; English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX CC polynucleotides encoding such proteins. The DPP peptides are useful for
XX CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX CC rejection and HIV (human immunodeficiency virus) infection. The present
XX CC Note: This sequence is stated to be the same as that shown as
XX CC SEQ ID NO: 4 in the sequence listing of the specification. However these
XX CC sequences differ.
SQ Sequence 847 AA:
QY 4 TGGPTADRGDAATDDPAARFOVOKHSMGILRSITIHGSRKYSGLIYKAPRDFQVOKTD 63
DB 10 SGVSPVQVAAAGMDDTAARFCVQKSHMDGILRSITIHGSRKYSGLIYKAPRDFQVOKTD 69
QY 64 ESGPHSHRLYYLGMYPGSRNSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 123
DB 70 ESGPHSHRLYYLGMYPGSRNSLLYSEIPKKVKREALLLSWKOMLDHFOATPHHGYSR 129
QY 124 EEBLLRRKRRLGVGITSYDFHSGSLFLFOANSFLHCRDGGKNGFMVSPMKPLEIKTQ 183
DB 130 EEBLLRRKRRLGVGITSYDFHSGSLFLFOANSFLHCRDGGKNGFMVSPMKPLEIKTQ 189
QY 184 CSGPRMDPKICPADPAFFSFINSNDLVANIENTGEERLFLFCHOGLSNVLDPPKSAGVAT 243
DB 190 CSGPRMDPKICPADPAFFSFINSNDLVANIENTGEERLFLFCHOGLSNVLDPPKSAGVAT 249
QY 244 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVYIHVSPALBERKTDY 303
DB 250 FVIOEEFDRFTGYWMCPTASWESSEGLKTLRLIYEEDSEVEVYIHVSPALBERKTDY 309
QY 304 RYPTGSKNPKIALKLAELFOTDSOGKITVTOEKELVOPFSSLPFKVYIARAGTTRGKY 363
DB 310 RYPTGSKNPKIALKLAELFOTDSOGKITVTOEKELVOPFSSLPFKVYIARAGTTRGKY 369
QY 364 AMAMFLDRPOOMLOLVLLPPALFIPSTENEBOQLASARAARVPRNQPYYVEEVTNWINV 423
DB 370 AMAMFLDRPOOMLOLVLLPPALFIPSTENEBOQLASARAARVPRNQPYYVEEVTNWINV 429
QY 424 HDIFPRPOSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSQGYDWSPEFGEDEFKCP 483
DB 430 HDIFPRPOSEGEDELCEFLRANECKTGFCCHLYKVTAVLKSQGYDWSPEFGEDEFKCP 489
QY 484 KEELALTSGEWEVLARHGSKIWNNEETKLYVFOGKTPTLEHNLVYVSYEAGETVRLTT 543

```

Db 490 KEEVALTSGEWEVLRSHGSKIMVNEQGTCLVYFGTKOTPLEHHLVYVESAGEIVRLTT 549
QY 544 PGFHSCHSCMSQNFDMFVSHYSVSSTPPCVNHYKLSGPRDDPLHKQPRWASMEASCPP 603
Db 550 LGFHSCHSCMSQSFDMFVSHYSVSSTPPCVNHYKLSGPRDDPLHKQPRWASMEASCPP 609
QY 604 DYVPEIFHFHTRSDVRLYGMITYKPKHALQPKKHPITVLFVYGGPOVOLVNNNSFKIGIKLR 663
Db 610 DYVPEIFHFHTRADVOLXGMITYKPKHPTLQPGKRPHTVLFVYGGPOVOLVNNNSFKIGIKLR 669
QY 664 LNTLASLGYAVVVIDGRSGCGRLRFEGALKNQMGVEIEDQVEGLQFAEKYGFIDLSR 723
Db 670 LNTLASLGYAVVVIDGRSGCGRLRFEGALKNQMGVEIEDQVEGLQFAEKYGFIDLSR 729
QY 724 VAHGMSTGGFSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERYMDVPENNQHGTEA 783
Db 730 VAHGMSTGGFSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERYMDVPENNQHGTEA 789
QY 784 GSVALLHVEKLEPNERNLLILHGFLDENVHFHTNPLVSQLIRACKPYOLOI 834
Db 790 GSVALLHVEKLEPNERNLLILHGFLDENVHFHTNPLVSQLIRACKPYOLOI 840

RESULT 10
ABG61605
ID ABG61605 standard; Protein: 832 AA.
XX

AC ABG61605:

DT 12-AUG-2002 (first entry)

DE Human DPRP-2 splice variant #4.

XX
KW Human: serine protease: dipeptidyl peptidase IV-related protein; DPRP;
KW DPPiV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinetic; reproductive disorder; inflammatory disorder;
KW metabolic disorder.

OS Homo sapiens.

PN W020023134-A2.

PD 18-APR-2002.

PE 12-OCT-2001; 2001WO-US31874.

PR 12-OCT-2000; 2000US-240117P.

PA (FERR) FERRING BV.

PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;

DR WPI: 2002-444178/47.

DR N-PSDB: ABK83336.

XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain

PS Disclosure; Page 85-87; 113pp; English.

XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPiV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,

CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.

XX
SQ Sequence 832 AA:

Query Match 86.7%; Score 4026; DB 23; Length 832;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 757; Conservative 6; Mismatches 16; Indels 20; Gaps 3;

QY 1 MATTGTPADGDAATATDPPARQOVOKHSDGLSIHSGSKTSGLIYNKAPHDFOVQ 60
Db 30 MATTGTPADGDAATATDPPARQOVOKHSDGLSIHSGSKTSGLIYNKAPHDFOVQ 89
QY 61 KTDGSPHSHRLYYLGMFYSGRENSLYSEIPIKRYKRFALLLSMKOMLDHFOATPHHGV 120
Db 90 KTDGSPHSHRLYYLGMFYSGRENSLYSEIPIKRYKRFALLLSMKOMLDHFOATPHHGV 149
QY 121 YSREELLREKRLGVFGITSYDFHSESGFLFQASNSLFHCROGKNGFVSPKPLEI 180
Db 150 YSREELLREKRLGVFGITSYDFHSESGFLFQASNSLFHCROGKNGFVSPKPLEI 209
QY 181 KTQCGPRMDPKICPADAPFSTINSDLWVANITGGERRLTFCHQGLSNVLDPKSAG 240
Db 210 KTQCGPRMDPKICPADAPFSTINSDLWVANITGGERRLTFCHQGLSNVLDPKSAG 269
QY 241 VATFVIOEEPRFTGYWMCPTASWEGSEGLKTLRLIYEEVSESEVYHVSPPALEEKT 300
Db 270 VATFVIOEEPRFTGYWMCPTASWEGSEGLKTLRLIYEEVSESEVYHVSPPALEEKT 329
QY 301 DSYRPRGSKNPKIALKLAEFYQDSQKIVSTQKELVOPFSSLFPRVEXIARAGWTRD 360
Db 330 DSYRPRGSKNPKIALKLAEFYQDSQKIVSTQKELVOPFSSLFPRVEXIARAGWTRD 389
QY 361 GRYAMAMLDPRQOMLOVLLPRLAFTSTENEEDRLASARAVRPNVYVVEEYVNW 420
Db 390 GRYAMAMLDPRQOMLOVLLPRLAFTSTENEEDRLASARAVRPNVYVVEEYVNW 449
QY 421 INVHDIFFYRPOSAGEDLCLFRANECKTGFCCHLYKVAVLKSQGYDMSSEPFSPCEDFEK 480
Db 450 INVHDIFFYRPOSAGEDLCLFRANECKTGFCCHLYKVAVLKSQGYDMSSEPFSPCEDFEK 509
QY 481 CPIKEEIALTSGEWEVLARHGSKIMVNEETKLVYFGTKDPRLEHHLVYVESAGEIVR 540
Db 510 CPIKEEIALTSGEWEVLARHGSKIMVNEETKLVYFGTKDPRLEHHLVYVESAGEIVR 569
QY 541 LTTPEFHSCHSCMSQNFDMFVSHYSVSSTPPCVNHYKLSGPRDDPLHKQPRWASMEAS 600
Db 570 LTTPEFHSCHSCMSQNFDMFVSHYSVSSTPPCVNHYKLSGPRDDPLHKQPRWASMEAS 629
QY 601 CPPDVVPEIFHFHTRSDVRLYGMITYKPKHALQPKKHPITVLFVYGGPOVOLVNNNSFKIG 660
Db 630 CPPDVVPEIFHFHTRSDVRLYGMITYKPKHALQPKKHPITVLFVYGGPOVOLVNNNSFKIG 689
QY 661 YLRNLTASLGYAVVVIDGRSGCGRLRFEGALKNQMGVEIEDQVEGLQFAEKYGFID 720
Db 690 YLRNLTASLGYAVVVIDGRSGCGRLRFEGALKNQMGVEIEDQVEGLQFAEKYGFID 749
QY 721 LSRVAIHGWSYGGFSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERYMDVPENNQHG 780
Db 750 LSRVAIHGWSYGGFSLMGLIHKPOVFKVALAGAPVYVMAVYDGYTERYMDVPENNQHG 789
QY 781 YEAGSVALLHVEKLEPNERN 799
Db 790 RKRALFPHKLRPLRTPDSPR 808

RESULT 11
ABG61606
ID ABG61606 standard; Protein: 832 AA.

XX ABG61606;
AC
XX 12-AUG-2002 (first entry)
DT
XX
DE Human DPPR-2 splice variant #5.
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinnesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.
XX
XX Homo sapiens.
OS
XX WO200231134-A2.
PN
XX 18-APR-2002.
PD
XX 12-OCT-2001; 2001WO-US31874.
PF
XX 12-OCT-2000; 2000US-240117P.
PR
XX (FERR) FERRING BV.
PA
XX Q1 S, Akinsanya KO, Riviere PJ, Junien J;
PI WPI; 2002-444178/47.
DR N-PSDB: ABK83337.
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
XX Disclosure; Page 88-90; 113pp; English.
PS
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinnesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
XX
SQ Sequence 832 AA;
Query Match 86.7%; Score 4026; DB 23; Length 832;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 757; Conservative 6; Mismatches 16; Indels 20; Gaps 3;

DB 210 KTQCSGPRMDPKICPADPAFFSFITNSDLWVANIETGEERLTFCHQGLSNVLDPPKSAG 269
QY 241 VATFVIOEEDRFTGYMWCPTASMESEGLTKLTILYEEVDESEVYIHWSPALAEKRT 300
DB 270 VATFVIOEEDRFTGYMWCPTASMESEGLTKLTILYEEVDESEVYIHWSPALAEKRT 329
QY 301 DSYRYPRTSKNPKIALKLAFFQDSOGKIVSTOEKELVOPFSSLPKVEYIARAGWTRD 360
DB 330 DSYRYPRTSKNPKIALKLAFFQDSOGKIVSTOEKELVOPFSSLPKVEYIARAGWTRD 389
QY 361 GKIVAMAFIDRPOQMLQVLPLPFLPTBTNEQRLASARAVRNQPVYVEEVYINW 420
DB 390 GKIVAMAFIDRPOQMLQVLPLPFLPTBTNEQRLASARAVRNQPVYVEEVYINW 449
QY 421 INVHDIFFPPQSGEDDELFLRANECKTGFCCHLYKTAVALKSOGYMSSEPFSGEDEPK 480
DB 450 INVHDIFFPPQSGEDDELFLRANECKTGFCCHLYKTAVALKSOGYMSSEPFSGEDEPK 509
QY 481 CPIKEEIALTSGEWEVLARHSGKIMVNEETKLVYFOGKTPTLEHHLYVVSYEAGETIVR 540
DB 510 CPIKEEIALTSGEWEVLARHSGKIMVNEETKLVYFOGKTPTLEHHLYVVSYEAGETIVR 569
QY 541 LTPPGFSHSCMSQNFDMFVSHSVSTPCVHYKLSGPDDELHKOPRFWASMEAS 600
DB 570 LTPPGFSHSCMSQNFDMFVSHSVSTPCVHYKLSGPDDELHKOPRFWASMEAS 629
QY 601 CPPDYVPEPIFHFTRSDVRLYGMIVKPHALQPKKHPTVLFYVGGPOVOLVNSPFGIK 660
DB 630 CPPDYVPEPIFHFTRSDVRLYGMIVKPHALQPKKHPTVLFYVGGPOVOLVNSPFGIK 689
QY 661 YLRNLTLASLGAAVVIDGRSCORGLFEGALKNQGOVEIEDOVGALQFAEKYFTID 720
DB 690 YLRNLTLASLGAAVVIDGRSCORGLFEGALKNQGOVEIEDOVGALQFAEKYFTID 749
QY 721 LSRVAIHGMSYSGFLSLMGLIHKPOVKVIAAGAPVYMAAYIDGYERYVADVENNQHG 780
DB 750 LSRVAIHGMSYSGFLSLMGLIHKPOVKVIAAGAPVYMAAYIDGYERYVADVENNQHG 789
QY 781 YEAGSVALLHEKLPNEPR 799
DB 790 RKRALFPKRLPRLPTDPSR 808
RESULT 12
ABG61609
ID ABG61609 standard; Protein: 819 AA.
XX
AC ABG61609;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPPR-2 splice variant #8.
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinnesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.
XX
XX Homo sapiens.
OS
XX WO200231134-A2.
PN
XX 18-APR-2002.
PD
XX 12-OCT-2001; 2001WO-US31874.
PF
XX 12-OCT-2000; 2000US-240117P.
PR
XX (FERR) FERRING BV.
PA

	QY	601 CPEDVYPPELFPHHTSTSDRLKGMITYKPALOPGKKHPTVLFPVGGPVOLVNNSFGIK 660
	Dd	617 CDPDVPPELFFHHRTSDDLRYLGMYTKPHALPPGGKHPTLVLPVYGGPVOLVNNSEFKIG 676
	OY	661 YLRNLNTLASGVAVVVVIDRSGSCORGLREPGALKNMOCVELEPOVESTOLFVAKEGYFD 720
	Dd	677 YLRNLNTLASGVAVVVVIDRSGSCOKLREPAGALKNMOGEIEOQELOGFPAEKGYEID 736
	OY	721 LSRVALHGMSYGGSFLSMGLIHKRPQVFKAIALGAEPVTWMAVDITGYTERVMADVENNQHG 780
	Dd	737 LSRAIHGNMYSYGFSLSMGILHKRPQVK-----AQPLATPR---LP-----G 776 : :
	OY	781 YEAGSVALLHEKLIPNEPNR 799 :
	Dd	777 RKRALPFPHKLPRLPTPSR 795 :
		RESULT 13
	ABG61610 ID	ABG61610 standard; Protein; 819 AA.
	XX AC	ABG61610;
	DT DT	12-AUG-2002 (first entry)
	DE DE	Human DPRP-2 splice variant #9.
	XX XX	Human: serine protease; dipeptidyl peptidase IV-related protein; DPPP; KW DPPIV, infection; human immunodeficiency virus; HIV-1; HIV-2; pain; KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke; KW heart failure; hypertension; urinary retention; osteoporosis; cancer; MW ulcer; allergy; cancer; psychotic disorder; neurological disorder; KM dyskinnesia; reproductive disorder; inflammatory disorder; KM metabolic disorder. XX Homo sapiens. OS WO200231134-A2. FN PN XX 16-Apr-2002. PD PD XX 12-Oct-2001, 2001WO-US31874. PF PF XX 12-Oct-2000; 2000US-240117P. PR PR XX (FEER) FERRING BV. PA PA XX Qi S, Akinsanya KO, Riviere PJ, Junien J; XX WPI; 2002-444178/47. DR DR N-PADB; ABX83341. XX XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding the proteins, useful for treating e.g. fungal, bacterial, protozoan and viral infections, cancers, allergies, neurological disorders, or pain -
	ES ES	Disclosure: Page 101-103; 113pp; English.
	CC CC	The present invention relates to the isolation of novel human serine proteases referred to as dipeptidyl peptidase IV (DPPIV)-related proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP) and nucleic acids encoding them are useful for treating infections such as fungal, bacterial, protozoan and viral infections, particularly infections caused by human immunodeficiency virus (HIV-1 or HIV-2), pain, diabetes, precocious puberty, infertility, obesity, anorexia, bulimia, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, stroke, ulcers, asthma, allergies, cancers, migraine, vomiting, psychotic and neurological disorders (e.g. anxiety, dementia, or schizophrenia), and dyskinesias. These may also be used in discovering therapeutic agents for the treatment of reproductive, inflammatory and

Dd	121	GLSNVLDDPKSAGATVIVIQEEDPRFICGMWCCPTASMSSEBGLKTLRLILEVDDESEVEN	180
Oy	288	IHWDSPALEEKRTDYSYRPTGSKNPKIALKLAEFQTDSGKITVSQEKELVOPFSSLEP	347
Dd	181	IHWSPALAEKKRTDYSYRPFGSKNPKIALKLKEFCQDMSGKITVSQEKELVPFSSLEP	240
Oy	348	KVEYIARAGWTRODKRYAMAMFLDRPOOMLOLVLLPRALFTPSNEBEOQLASRAAPRVN	407
Dd	241	KVEYIARAGWTRDCKRYAMAMFLDRPOOMLOLVLLPRALFTPSNEBEOQLASARAAPRVN	300
Oy	408	OPYVVYEETVTVMWJINVHDIFYPFQSGBDELCTLRANECKTGTCCHLYKTATVALKSOGYD	467
Dd	301	OPYVVYEETVTVMWJINVHDIFFPFQSGEBDELCTLRANECKTGTCCHLYKTATVALKSOGYD	360
Oy	468	WSEPFSPGEDEFCFKPIKEEIALTSGEMEWLARHGSKIWMEETKLKYVFQGTCDTPLENHL	527
Dd	361	WSEPFSPGEEG-----EGSLTNA-----IWNVEETKLKYVFQGTCDTPLENHL	401
Oy	528	YVYSIEAAGEIVLRTLTPGFHSHCMSMQNFDMFYSHSYSSYSTPPCVNHYYKLSGGDDDLRK	587
Dd	402	YVYSIEAAGEIVLRTLTPGFHSHCMSMQNFDMFYSHSYSSYSTPPCVNHYYKLSGDDDDLHK	461
Oy	588	QPRFMASMMELAASCPDPVPEIPFHPTRSDFLYGMYIKPRHALOPKKHPVLFVYGGP	647
Dd	462	QPRFMASMMELA-----KIHFHTRSDFVRLYGMTIKPRHALQPEKKHPVLFVYGGP	512
Oy	648	QVOLVNNSFKGIKYLRLNTLASLGAYAVVVIDGRSCCGRGILRFEGALKNMQGOVEIEDOYE	707
Dd	513	QVOLVNNSFKGITKYLRLNTLASLGAYAVVVIDGRSCCGRGILRFEGALKNMQGOVEIEDOYE	572
Oy	708	GLQFYAAEKYGFIDLSRAVAIHGWSYGGFLSLMGLIHKPQVERKVALIAGAPVTWMAYPDGYT	767
Dd	573	GLOFYAAEKYGFIDLSRAVAIHGWSYGGFLSLMGLIHKPQVERKVALIAGAPVTWMAYPDGYT	632
Oy	768	EERYMDVPNNNGHYEAGSVALLHYEKLRENERPLLILHGFIDENVHEPHTPLYSOLIRAG	827
Dd	633	EERYMDVPNNNGHYEAGSVALLHYEKLRENERPLLILHGFIDENVHEPHTPLYSOLIRAG	692
Oy	828	KPYQL-----QIYPNERHSTRICESGHEHYEVTLTFHFOEYL	863
Dd	693	KPYQLVALPVPVSQIYPNERHSTRICESGHEHYEVTLTFHFOEYL	737
<hr/>			
RESULT 15			
ID	AAM40510	standard; Protein: 683 AA.	
XX	AAM40510:		
AC	22-OCT-2001	(first entry)	
DT	Human polypeptide SEQ ID NO 5441.		
DE	Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer		
XX	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
OS	Homo sapiens.		
PN	MO200153312-A1.		
XX	26-JUL-2001.		
PE	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA44213) with neurotrophic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Query Match	76.28;	Score 3539.5;	DB 22;	Length 683;
-------------	--------	---------------	--------	-------------

QY	162	CDDGKNGFVMSYMKPRLKIKQSCSPRRDPICADPFFSFJINNSDLWANIETGERR	221
Db	1	CRDGGKNGFVMSYMKPRLKIKQSCSPRRDPICADPFFSFJINNSDLWANIETGERR	60
QY	222	LTFCHQGISNVLDDPKSGVATFYQEEFDRFTGYWMCPTASWEGSBGLKTLRIIYEVD	281
Db	61	LTFCHQGISNVLDDPKSGVATFYQEEFDRFTGYWMCPTASWEGSBGLKTLRIIYEVD	120
QY	282	ESEVYHVHPSPALEREKTDYRYRPRGSKNPKITALKLAETOSOCKIYSTOKELVOP	341
Db	121	ESEVYHVHPSPALEREKTDYRYRPRGSKNPKITALKLAETOSOCKIYSTOKELVOP	180
QY	342	FSSLEPKVEYLIRAGWTRDKGAKAMAFDRQOMJOLVLLPRLPISTENEBORLASAR	401
Db	181	FSSLEPKVEYLIRAGWTRDKGAKAMAFDRQOMJOLVLLPRLPISTENEBORLASAR	240
QY	402	AVPRNVOPRYUYEEVYNWNVINHDIYFPPPOSEGEDELCLIRANECKTGFCNLYKVTAVL	461
Db	241	AVPRNVOPRYUYEEVYNWNVINHDIYFPPPOSEGEDELCLIRANECKTGFCNLYKVTAVL	300
QY	462	KSQGTDWSEPRSPRGDEKCRPKIEALATSEBWEVLAAHSGKIYNVNETKKIUYPOGKJOT	521
Db	301	KSQGTDWSEPRSPRGDEKCRPKIEALATSEBWEVLAAHSGKIYNVNETKKIUYPOGKJOT	341
QY	522	PLEHNLUVYSEAEIYRLTTPGFSHSGSCSOMDFMVSHTSSVTPRCVNYVLLSGPD	581
Db	342	PLEHNLUVYSEAEIYRLTTPGFSHSGSCSOMDNMVSHTSSVTPRCVNYVLLSGPD	401
QY	582	DDPLKORPFWASMKMEASCPEDYVPRLEIHFHTRSDVRLXGMTYKRDHALOPGKKNPVL	641
Db	402	DDPLKORPFWASMKMEAL-----KIFPHHTSSVDRLXGMTYKRDHALOPGKKNPVL	452

Qy 642 FVYGGPOVQVNNNSFKIKYLRNLTLASLGAVVVIDGRGSCQGRGLRREGALKNOMGOVE 701
Db 453 FVYGGPOVQVNNNSFKIKYLRNLTLASLGAVVVIDGRGSCQGRGLRREGALKNOMGOVE 512
Qy 702 IEDOVEGLQFAEKYGFIDLSRVAIHGWSYGGFSLMGLIHKPOVFKVAIAGAPVTWMA 761
Db 513 IEDOVEGLQFAEKYGFIDLSRVAIHGWSYGGFSLMGLIHKPOVFKVAIAGAPVTWMA 572
Qy 762 YDTGYTERYMDVPENNONGYEAGSVALHVEKLPNEPNRLLILHGFLEDNVHFFHTNFLVS 821
Db 573 YDTGYTERYMDVPENNONGYEAGSVALHVEKLPNEPNRLLILHGFLEDNVHFFHTNFLVS 632
Qy 822 QLIRAGKPYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQEYL 863
Db 633 QLIRAGKPYQLQVALPPVSPQIYPNERHSIRCPESGEHYEVTLLHFLQEYL 683

Search completed: December 12, 2002, 12:04:24
Job time : 37 secs

Db 819 TACCACCCGAGAGCGGAAACATTCCTGTTTCAAGCTGGACGCGCATTTACCATATCAAA 878
 Qy 500 GAGCGCGCAAGAACGGCTTCATGCTGTCCCTATGAAACCGCTGGAATCAAGCCAG 559
 Db 879 GATGGAGGACCAATGATGATTTACACACAGCGCTTGGCGCCCAATTTTGTGAAATCTAGT 938
 Qy 560 TGCTCAGGGCCCGATGAGCCCAAAATTCGCCCTGGCAGCCCTGCTTCTTCCTTC 619
 Db 939 TGTCGCATATATACGATGATGCCAAATATATGCTGCTGATCCGAGATGGATACCTTTC 998
 Qy 620 ATCAATTAACAGCAGCCTGTGGTGGCCACATCGACAGAGCGAGAGCGCGGCTGAC 679
 Db 999 ATTCAACAGCAATGATATTTGGATATCAAACTCGTAACAGGAGAGAGCGAGATCA 1058
 Qy 680 TTTCGCCACCAAGGTTTTCATGCTCTGATGACCCCAAGCTGCGGGGTGTGCCACC 739
 Db 1059 TACGTACCAATGAGCTGAGCAACATGGAAGAGATCCCAATCAGCTGGTGGTCCACC 1118
 Qy 740 TTGCTCATACAGAAAGATTCGACCGCTTCACCTGGTGGTGGTGGTGGTGGTGGTGG 799
 Db 1119 TTTGCTCTTCAAGAAAGATTTGACAGATACCTGCTACTGCTGCTGCTGCTGCTGCTG 1178
 Qy 800 TGGAGAGGTTGAGAGGCGCTCAAGACGCTGCAATCTGTATGAGAAAGTGAATGATCC 859
 Db 1179 AG---AACCTCAGTGTGTGTAATTTCTGAAATTTCTGTAAGAAATGATGATCT 1235
 Qy 860 GAGGTGAGGTCATTCACGTCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 919
 Db 1236 GAGGTGAGGATTAATTCATGTTACGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
 Qy 920 CGGTATCCCGAGAGAGGAGCAAGATCCCAAGATTCGCTGAACTGGCTGATGTTCCAG 979
 Db 1296 CGTATATCCCAAAACAGCAGCAAGCAAAAGCTCACTTCAGATGTCCGAGATGTT 1355
 Qy 980 ACTGACAGCCAGGGCAGATGCTCGACCCAGAGAAAGAGAGCTGTCAGCCCTTCAGC 1039
 Db 1356 GTTATGCTGACGAGGAGATTAATGATGTCATGATTAAGAACTGTTCAACCTTTCAG 1415
 Qy 1040 TCGCTGTTCCCGAAGGTGAGATACATGCGAGGCGCGGCTGACCCCGGATGCGAAATAC 1099
 Db 1416 ATTCGTTGAGGAGGATTAATTAATTCAGAGCGGAGTACCTCAGAGGAAATAT 1475
 Qy 1100 GCTGCGGCGATGTTCTCTGAGACGGCGCCAGAGTGGCTCAGCTGCTCTCTCCCGG 1159
 Db 1476 GCTGCGGCTGATCTACTAGACCGTTCCAGACTCAGCTGATGTTCTGCTCTCCCT 1535
 Qy 1160 GCCCTGTTCAATCCGAGCAGAGATGAGAGCAGCGGCTAGCTGCGCAGAGCTGTC 1219
 Db 1536 GAGTTATTCATCCAGTGAAGATGATGCTATGACAGAGAGACTTATGAGTCAAGTT 1595
 Qy 1220 CCCAGGAATTCAGGCGGTATGCTGATGAGAGGAGTCCACCAAGCTGATCAATGTT 1279
 Db 1536 CCGTACTGTGACACCACTATATCTATGAGAAACAGACGATGTGATTAATATC 1655
 Qy 1280 CATGACATCTTCTATCCCTCCATCAGAGGAGAGAGAGAGAGCTGCTTCTCCGC 1339
 Db 1656 CACGATATTTTCAATGTTTCTCTCAACTCA---TGAAGATGAATTTGAGTTTATTTT 1712
 Qy 1340 GCCATGATGACAGAGCGGCTTCTGCAATTTGTAACAAGTACCGCGTTTAAATCC 1399
 Db 1713 GCTCTGATGACAAACAGGTTTCTGCTATGATTAATAACATCCATTTTAAAGAG 1772
 Qy 1400 CAGGGCTGCAATGAGAGAGCCCTTCAGCCCGGGAAGATGAATTAAGTCCCATTT 1459
 Db 1773 AGCAATATTAACGGTCCAGTGTGACTACTGCTGCCCAAGGATTTAAAGTCTATC 1832
 Qy 1460 AAGGAAGATGCTCTGACAGCGGATGAGAGTGTGCGAGGACAGCGCTCAAG 1519
 Db 1833 AAAGAGAAATTAACATTAACAGAGTGTGATGAGAGTACTTGGCGGACATGATCTAAT 1892
 Qy 1520 ATCTGGCTCAATGAGAGACCAAGCTGCTGATCTTCCAGGGCACCAGAGCGCGCTG 1579
 1111111 1111 1 11111111111 1111111 11 11

Db 1893 ATCTGGTTGATGAAGCCAGAAAGCTGCTACTTTGAAGGACCAACAGCTCTCTTG 1952
 Qy 1580 GAGCACACCTCTACGTGTGATGATGAGCGCGCGGAGATGATGCTCTACCCAG 1639
 Db 1933 GAGCATACCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2012
 Qy 1640 CCGGCTCTCTCCATGATGCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1699
 Db 2013 GGTGCTACTGACACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2072
 Qy 1700 AGCAGCTGAGACAGCGCGCTCTGCTGACGCTTCAAGAGTGAAGCGCGCCGACAGCAG 1759
 Db 2073 AGCAACAGAAAGAAATCCACACTGTGTCTCTCAAACTCTCAAGCTGAGAGATAC 2132
 Qy 1760 CCGCTGACAAAGACCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1819
 Db 2133 CCGATTCTAAACAAAGAAATTTTGGGCGACCATTTTGGATTCAGAGGCTCTTCTCT 2192
 Qy 1820 GATTATGTTCTCCAGAGATCTTCCATTTCCACAGCGCGCTGATGCTGCGCTTACGCG 1879
 Db 2193 GACTACACCCCTCCAGAAATTTTCTTTTGAAGTACTACTGATTTACACTGTATGSA 2252
 Qy 1880 ATGATCTACAGCCCGACCGCTTTCAGCCAGGAGAAACACCCACCGCTCTTGTGA 1939
 Db 2253 ATGTTGATTAAGCCTCATGACTTACACCTGGAAGAAATACCCCATGTGTTATCATTA 2312
 Qy 1940 TATGAGGCGCCCGAGTGTGATGATTAATCTCTTCAAGAGCTCAATGATCTGCTG 1999
 Db 2313 TATGTTGTCTCCAGAGTGTGATGATTAATCGGTTTAAAGAGTCAAGATTTCCGC 2372
 Qy 2000 CTCAACACACTGCGCTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2059
 Db 2373 CTGAACACCTCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2432
 Qy 2060 CAGCGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2119
 Db 2433 CACCGAGACTTAATTTGAAGGCGCTTTAAATTAATTAATTAATTAATTAATTAAT 2492
 Qy 2120 GACCGAGTGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2179
 Db 2493 GATCAAGTGAAGAGCTCAGTACCTGATGATGATGATGATGATGATGATGATGATGAT 2552
 Qy 2180 GTTGCATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2239
 Db 2553 GTGGGATTCACAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2612
 Qy 2240 CCGCAGGCTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2299
 Db 2613 TCGGATATCTTCCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2672
 Qy 2300 ACAGGCTACAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2359
 Db 2673 ACAGGATACAGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2732
 Qy 2360 GATTCCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2419
 Db 2733 GATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2792
 Qy 2420 CAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2479
 Db 2793 CATGGTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2852
 Qy 2480 ATCGAGCAGGGAACCTTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2539
 Db 2853 GTGAGGCTGGAAGCCATTAAGTACTTACAGATCTTCTCAGGAGAGCGACGATCAGA 2912
 Qy 2540 TGCCCGAGTGGGCGAGCAGCTATGATGATGATGATGATGATGATGATGATGATG 2599
 Db 2913 GTCTGATGCTGAGAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 2972
 Qy 2600 TGAGCG 2605
 Db 2973 GGATCG 2978

RESULT 2	BM553230	1137 bp	mRNA	linear	EST 20-FEB-2002
LOCUS	BM553230				
DEFINITION	AGNCOURT_6572641 NIH_MGC_41 Homo sapiens				
ACCESSION	BM553230				
VERSION	BM553230.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1137)				
TITLE	NIH-MGC http://mhc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: c9abcs@remail.nih.gov Tissue procurement: DCTD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov plate: L1CM1968 row: j column: 12 High quality sequence stop: 672.				
FEATURES	Location/Qualifiers				
SOURCE	1..1137 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5467115" /clone_id="NIH_MGC_41" /tissue_type="amelanotic melanoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(g). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."				
BASE COUNT	253 a 346 c 320 g 216 t				
ORIGIN	2 others				
Query Match	34.3%; Score 897.2; DB 13; Length 1137;				
Best Local Similarity	98.5%; Pred. No. 9.4e-186;				
Matches	947; Conservative 0; Mismatches 9; Indels 5; Gaps 4;				
OY	348 AGGCACGCCCCCAGCATGGGGCTTACTCTCGGGAGAGAGAGCTGCTGAGGAGCGGAAC	407			
Db	1 AGGCACGCCCCCAGCATGGGGCTTACTCTCGGGAGAGAGAGCTGCTGAGGAGCGGAAC	60			
OY	408 GCTGGGGGGCTTTCGGCATCACTCTCAAGCACTTCCACAGGAGAGTGCGCTTCCTCT	467			
Db	61 GCTGGGGGGCTTTCGGCATCACTCTCAAGCACTTCCACAGGAGAGTGCGCTTCCTCTCT	120			
OY	468 TTCAGGCCAGCAACAGCTCTTCACATGTGCGAGCGGCGGAAGAAGCGCTTCATGTGT	527			
Db	121 TTCAGGCCAGCAACAGCTCTTCACATGTGCGAGCGGCGGAAGAAGCGCTTCATGTGT	180			
OY	528 CCCCATCAAAACCGTGGAAATCAAGACCCAGTGCTCAGGGCCCCGGATGTGAACCCCAAAA	587			
Db	181 CCCCATCAAAACCGTGGAAATCAAGACCCAGTGCTCAGGGCCCCGGATGTGAACCCCAAAA	240			
OY	588 TCTGCCCCCGAGACCTTCCTCTTCCTTCATCAATAAAGAGCACTGTGGGTGGCA	647			
Db	241 TCTGCCCCCGAGACCTTCCTCTTCCTTCATCAATAAAGAGCACTGTGGGTGGCA	300			
OY	648 ACATCGAGACAGGCGGAGAGCGGGGCTGACCTTCGACCACCAAGTTTATCAATGTGC	707			

Db	301	ACATGAGACAGGGCAGGAGCGGGCGGTGACCTCTCTGCCACCAAGTTTATCCAAATGTC	360
OY	708	TGATGACCCCAAGTCTCTCGGGTGTGGCCACCCTTGCTATACAGAGAGATTGACCGCT	767
Db	361	TGGATGACCCCAAGTCTCTCGGGTGTGGCCACCCTTGCTATACAGAGAGATTGACCGCT	420
OY	768	TCACGGGTACTGGTGGTGGCCACAGCGCTCTGGGAAGTTCCAGAGGGCTCAAGAGCG	827
Db	421	TCACGGGTACTGGTGGTGGCCACAGCGCTCTGGGAAGTTCCAGAGGGCTCAAGAGCG	480
OY	828	TGCGAATCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAAGTCATTCACGTCCCTCTC	887
Db	481	TGCGAATCTGTATGAGGAAGTCGATGAGTCCGAGGTGGAAGTCATTCACGTCCCTCTC	540
OY	888	CTGGCTTGAGGAAGGAAGACGACCTCTGTCGATACCCAGGACAGGCAAGAAATC	947
Db	541	CTGGCTTGAGGAAGGAAGACGACCTCTGTCGATACCCAGGACAGGCAAGAAATC	600
OY	948	CCAAATTTGCTTGAACCTGAGTCCAGTCCAGACTGACAGCCAGGCGAAGATCTCTCGA	1007
Db	601	CCAAATTTGCTTGAACCTGAGTCCAGTCCAGACTGACAGCCAGGCGAAGATCTCTCGA	660
OY	1008	CCCAGGAAGAAGAGCTGTGTGACGCCCTTACACTGCTCTTCCGAAGTGGAGTACATCG	1067
Db	661	CCCAGGAAGAAGAGCTGTGTGACGCCCTTACACTGCTCTTCCGAAGTGGAGTACATCG	720
OY	1068	CCAGGCGCGGGTGGACCCGGGATGGCAATACGCTGGGCGCATTTCTTGACCGGCCCC	1127
Db	721	CCAGGCGCGGGTGGACCCGGGATGGCAATACGCTGGGCGCATTTCTTGACCGGCCCC	780
OY	1128	AGCACTGCTCCAGTCTGCT	1187
Db	781	AGCACTGCTCCAGTCTGCT	839
OY	1188	AGGACACACGGCTAGCCCTGTCGACAGAGCTGTCCCGAGAAATGTCCAGCCGATGGGATG	1247
Db	840	AGGACACACGGCTAGCCCTGTCGACAGAGCTGTCCCGAGAAATGTCCAGCCGATGGGATG	899
OY	1248	ACGA-GGAGGTACCAACGCTCT-GGATCAATGTTTCATG-ACATCTCTATCCCTTCCC	1303
Db	900	ACGAGGAGGTACCAACGCTCTGGGATCAATGTTTCATGAGCAATCTTCTATCCCTTCCC	959
OY	1304	C 1304	
Db	960	C 960	

|||||
Db 541 GTCTCTCCCTCCCGCCGCTGTCTATCCGAGACAGAGATGAGAGACGCGCTACCC 600
Qy 1205 TCTCCAGAGAGCTGTCTCCAGAGATGTCCAGCCGATGTGCTGACGAGAGAGTCAACAC 1264
Db 601 TCTCCAGAGAGCTGTCTCCAGAGATGTCCAGCCGATGTGCTGACGAGAGAGTCAACAC 660
Qy 1265 GTCTGATCAATGTTCATGATGATCTTCTATCCCTTCCCAATCAG-AGGAGAGAGACGA 1323
Db 661 GTCTGATCAATGTTCATGATGATCTTCTATCCCTTCCCAATCAGAGAGAGAGACGA 720
Qy 1324 GCTGTGCTTCTCCGCGCCCAATGATGACAGCGGCTTCTGCAATTTGTACAAAGTCA 1383
Db 721 GCTGTGCTTCTCCGCGCCCAATGATGACAGCGGCTTCTGCAATTTGTACAAAGTCA 780
Qy 1384 CGCGCTTTAAATCCAGGCTACGATGAGTGAAGCCCTTCAAGCCCGGAGAAATGA 1443
Db 781 CGCGCTTTT-ANATCCAGGCTACCATTCGAGTGA-CCCTTCCGCGCCGAGAAATGA 838
Qy 1444 ATTTAAGTCCCATTAAGAGAGATGCTGTGACACGCGGTGA--TGGAGGCTTT 1500
Db 839 ATTTAAGTCCCATTTTAAAGAGAGATGCTGTGACACGCGGAGAAATGGAGAGTTT 898
Qy 1501 GCGGAGGACGCGCTCCAGATCTGGGTCAATG 1532
Db 899 GCGGAGGACGCGCTCCAAATTTGGGTCAAG 930

RESULT 6
B0678015 982 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8034317 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213986
DEFINITION 5', mRNA sequence.
ACCESSION B0678015
VERSION B0678015.1 GI:21790694
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 982)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNML Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.lnl.gov
Plate: LCM2380 row: 3 column: 03
High quality sequence stop: 683.
Location/Qualifiers
1. 982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6213986"
/clone_1lb="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 202 a 310 c 290 g 179 t 1 others
ORIGIN

Query Match 31.9%; Score 834; DB 14; Length 982;
Best Local Similarity 96.7%; Pred. No. 6e-172;
Matches 893; Conservative 0; Mismatches 26; Indels 4; Gaps 4;
Qy 1473 CTCTGACCAAGGCTGAATGGAGGTTTGGGAGGACAGGCTCCAGATCTGGGTCAATG 1532
Db 1 CTCTGACCAAGGCTGAATGGAGGTTTGGGAGGACAGGCTCCAGATCTGGGTCAATG 60
Qy 1533 AGGAGACCAAGCTGTGTACTTCCAGGGCACCAAGAGACAGCGCTGGAGACCACTCT 1592
Db 61 AGGAGACCAAGCTGTGTACTTCCAGGGCACCAAGAGACAGCGCTGGAGACCACTCT 120
Qy 1593 ACGTGTAGCTATGAGGCGGCGCGGAGATGATGATGATGATGATGATGATGATGATGAT 1652
Db 121 ACGTGTAGCTATGAGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 180
Qy 1653 ATAGCTCTCCATGAGGACGAGAACTTCCAGATGTTGTTGAGGCTACAGCAGCTGAGCA 1712
Db 181 ATAGCTCTCCATGAGGACGAGAACTTCCAGATGTTGTTGAGGCTACAGCAGCTGAGCA 240
Qy 1713 CGCGGCTTGTGCTGACGCTTACAAAGCTGAGCGGCGCCGAGAGACCCCTGACAAAGC 1772
Db 241 CGCGGCTTGTGCTGACGCTTACAAAGCTGAGCGGCGCCGAGAGACCCCTGACAAAGC 300
Qy 1773 AGCCCGCTTGTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1832
Db 301 AGCCCGCTTGTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 1833 CAGAGATCTTCATTTCCACAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1892
Db 361 CAGAGATCTTCATTTCCACAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 1893 CCCACGCTTGCAGGCGAGGAGAAAGACCCACCGCTCTTTGTATATGAGGCGCC 1952
Db 421 CCCACGCTTGCAGGCGAGGAGAAAGACCCACCGCTCTTTGTATATGAGGCGCC 480
Qy 1953 AGTGTGACCTGTGATTAATTCCTTCAGAAAGATCAATCACTGTGGGCTCAACACATGG 2012
Db 481 AGTGTGACCTGTGATTAATTCCTTCAGAAAGATCAATCACTGTGGGCTCAACACATGG 540
Qy 2013 CCTCCCTGGGTACGCGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2072
Db 541 CCTCCCTGGGTACGCGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Qy 2073 GATTGGAAGGGCCCTGAATAAATGAGGCGGAGATGAGAGAGAGAGAGAGAGAGAGAG 2132
Db 601 GATTGGAAGGGCCCTGAATAAATGAGGCGGAGATGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 2133 GCTTCGATGTTGTGCGGAGAAAGTATGGCTTCATGACCTGACCGAGATGGCATTCATG 2192
Db 661 GCTTCGATGTTGTGCGGAGAAAGTATGGCTTCATGACCTGACCGAGATGGCATTCATG 720
Qy 2193 GCTGTCTTACGAGGGGCTTCTCTCTGCTCATGAGGGCTATCCACAGCCGAGGTTC 2252
Db 721 GCTGTCTTACGAGGGGCTTCTCTCTGCTCATGAGGGCTATCC -CAAGCCAGAGTTC 779
Qy 2253 AGGTGGCCATGCGGGGTG-CCCGGTACACGCTTGATGAGGCTTACGACAGAGTACT 2311
Db 780 AGGTGGCCATGCGGGGTGCGGCGGCTCCCGGCTACCGCTGATGGCTTACGACAGAGTACT 839
Qy 2312 GAGC-GCTACATGAGGCTCCCTGAGACACAGCAGCGGTATGAGGGGGG-TTCGAGG 2369
Db 840 GAGCAGGCTTACATGAGGCTCCCTGAGACACAGCAGCGGTATGAGGGGGGTTTCGAGG 899
Qy 2370 CCTCTGACGCTGAGAGAGCTGCC 2392
Db 900 CCTCTGACGAGGAGAGAGAGCTGCC 922

RESULT 7
B0897707 880 bp mRNA linear EST 16-AUG-2002
LOCUS

DEFINITION AGENCOURT_8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208407
5', mRNA sequence.

ACCESSION B0897707
VERSION B0897707.1 GI:22289721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 880)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Plate: LNCM2366 row: a column: 16
High quality sequence stop: 704.

FEATURES
source
1. 880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6208407"
/clone_1ib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 194 a 260 c 249 g 176 t 1 others
ORIGIN

Query Match 30.8%; Score 805.2; DB 14; Length 880;
Best Local Similarity 97.9%; Pred. No. 1.2e-165;
Matches 848; Conservative 0; Mismatches 13; Indels 5; Gaps 3;

QY 605 GCCTTCTTCCTTCATCATACAGAGGAGCTGTGGTGGCCAAATCGAGACAGGGGAG 664
|||||
Db 1 GCCTTCTTCCTTCATCATACAGAGGAGCTGTGGTGGCCAAATCGAGACAGGGGAG 60
|||||

QY 665 GAGCGGGGCTGACCTCTGCGACCAAGTTTATCCATGCTCGATGAGACCCCAAGTCT 724
|||||
Db 61 GAGCGGGGCTGACCTCTGCGACCAAGTTTATCCATGCTCGATGAGACCCCAAGTCT 120
|||||

QY 725 GCGGGTGTGGCCACTTCGTATACAGAGAGTTTCGACCGCTTCACTGGGTACTGTGG 784
|||||
Db 121 GCGGGTGTGGCCACTTCGTATACAGAGAGTTTCGACCGCTTCACTGGGTACTGTGG 180
|||||

QY 785 TGGCCCAACACCTTCCTGGGAAGTTTCAGAGGGCTCAAGAGCTGGCAATCTCTATGAG 844
|||||
Db 181 TGGCCCAACACCTTCCTGGGAAGTTTCAGAGGGCTCAAGAGCTGGCAATCTCTATGAG 240
|||||

QY 845 GAAGTCATGAGTCCGAGGAGGAGGATTCAGCTCCCTCTCTGGGTAGAGAAAGG 904
|||||
Db 241 GAAGTCATGAGTCCGAGGAGGAGGATTCAGCTCCCTCTCTGGGTAGAGAAAGG 300
|||||

QY 905 AAGACGACTCGTATCGGTACCCAGAGACAGAGCAAGAAATCCCAAGATTGCTTGA 964
|||||
Db 301 AAGACGACTCGTATCGGTACCCAGAGACAGAGCAAGAAATCCCAAGATTGCTTGA 360
|||||

QY 965 CTGGCTGAGTTCAGACTGACAGCGGAGGAGAGATGCTCTGACCCAGAGAAAGAGCTG 1024
|||||

Db 361 CTGGCTGAGTTCAGACTGACAGCGGAGGCAAGATCTCTGACCCAGAGAAAGAGCTG 420
|||||

QY 1025 GTGAGCCCTTCACCTGCTGCTTCCCGAGAGTGAATACATCGCCAGGCGGGGTGAGC 1084
|||||

Db 421 GTGAGCCCTTCACCTGCTGCTTCCCGAGAGTGAATACATCGCCAGGCGGGGTGAGC 480
|||||

QY 1085 CGGATGGCAAAATACGCGCTGGCCATGTCCTGAGCCGCCAGCAGTGGCTTCACCTC 1144
|||||

Db 481 CGGATGGCAAAATACGCGCTGGCCATGTCCTGAGCCGCCAGCAGTGGCTTCACCTC 540
|||||

QY 1145 GTCTCTCTCCCGCGGCGGCTTTCATCCCGAGACACAGAAATGAGAGCAGCGGCTAGC 1204
|||||

Db 541 GTCTCTCTCCCGCGGCGGCTTTCATCCCGAGACACAGAAATGAGAGCAGCGGCTAGC 600
|||||

QY 1205 TCTGCCAGACTGTGCCAGAGATGTCACGCCGATGTGTGATGAGAGAGTACCAAC 1264
|||||

Db 601 TCTGCCAGACTGTGCCAGAGATGTCACGCCGATGTGTGATGAGAGAGTACCAAC 660
|||||

QY 1265 GTCTGATCAATGTTTCATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAG 1324
|||||

Db 661 GTCTGATCAATGTTTCATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAGAG 720
|||||

QY 1325 CTCTGCTTTCTCCCGCCCATGATGACAGACCGGCTTTCCTCCCTTCTCAAGTCAAC 1384
|||||

Db 721 CTCTGCTTTCTCCCGCCCATGATGACAGACCGGCTTTCCTCCCTTCTCAAGTCAAC 780
|||||

QY 1385 G-CGCTTTTAAATCCAGAGG--CTACGATTGAGTGAAGCC--TTACGCCCGGGGAG 1439
|||||

Db 781 GCCCGTTTAAATCCAGAGGCGCTACGATTTGAGTGAAGCCCTTTCAGCCCGGGGAG 840
|||||

QY 1440 ATGAATTTAAGTGGCCCATTAAGAA 1465
|||||

Db 841 ATGAATTTAAGTGGCCCATTAAGAA 866
|||||

RESULT 8
B0899360 879 bp mRNA linear EST 16-AUG-2002
LOCUS B0899360
DEFINITION AGENCOURT_8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208551
5', mRNA sequence.
ACCESSION B0899360
VERSION B0899360.1 GI:22281374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 879)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Plate: LNCM2366 row: g column: 16
High quality sequence stop: 709.

FEATURES
source
1. 879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6208551"
/clone_1ib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 194 a 256 c 252 g 177 t
ORIGIN

Query Match 30.0%; Score 785.8; DB 14; Length 879;

Best Local Similarity 99.6%; Pred. No. 2e-161;

Matches 798; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 605 GCCTCTCTCTCTCATATATACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 664
|||||
Db 1 GCCTCTCTCTCTCATATATACAGCAGCTGTGTGGCCCAACATCGAGACAGCGGAG 60
|||||
OY 665 GAGCGCGGCTGACCTCTTGTCCACCAAGTTTATCATATGTCCTGATGACCCCAAGTCT 724
|||||
Db 61 GAGCGCGGCTGACCTCTTGTCCACCAAGTTTATCATATGTCCTGATGACCCCAAGTCT 120
|||||
OY 725 GCGGCTGTGGCCACCTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 784
|||||
Db 121 GCGGCTGTGGCCACCTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 180
|||||
OY 785 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 844
|||||
Db 181 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 240
|||||
OY 845 GAAATCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAAAG 904
|||||
Db 241 GAAATCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAAAG 300
|||||
OY 905 AAGAGGAGCTCGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCTTGA 964
|||||
Db 301 AAGAGGAGCTCGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCTTGA 360
|||||
OY 965 CTGGGTGATTCACAGCTGACAGCAGGAGCAAGATCGTCTGACCCAGAGAGAGAGAGCTG 1024
|||||
Db 361 CTGGGTGATTCACAGCTGACAGCAGGAGCAAGATCGTCTGACCCAGAGAGAGAGAGCTG 420
|||||
OY 1025 GTGACAGCCTTCAGCTCGCTGTTCCTCCGAAGGTGAGTACATCGCAGGCGGCTGAGAC 1084
|||||
Db 421 GTGACAGCCTTCAGCTCGCTGTTCCTCCGAAGGTGAGTACATCGCAGGCGGCTGAGAC 480
|||||
OY 1085 CGGATGAGCAATATACGCTGTGGGCAATGTTCTTGACACCGGCCCCACAGTGTCTCAAGCTC 1144
|||||
Db 481 CGGATGAGCAATATACGCTGTGGGCAATGTTCTTGACACCGGCCCCACAGTGTCTCAAGCTC 540
|||||
OY 1145 GTGCTCTCTCCCGCGGCGCTGTTCATCCGAGACAGAGAAATGAGAGAGCGGCTAGCC 1204
|||||
Db 541 GTGCTCTCTCCCGCGGCGCTGTTCATCCGAGACAGAGAAATGAGAGAGCGGCTAGCC 600
|||||
OY 1205 TCTGCGAAGCTGTCCCGAGAAATGTCAGCCGTATGTGTGTCAGAGAGGTTCACCAAC 1264
|||||
Db 601 TCTGCGAAGCTGTCCCGAGAAATGTCAGCCGTATGTGTGTCAGAGAGGTTCACCAAC 660
|||||
OY 1265 GTGTGATCAATGTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGAGAGAG 1324
|||||
Db 661 GTGTGATCAATGTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGAGAGAG 720
|||||
OY 1325 CTCTGCTTCTCCCGGCAATGAAATGACAGCGGCTTCTGCAATTTGTAAATGTCAC 1384
|||||
Db 721 CTCTGCTTCTCCCGGCAATGAAATGACAGCGGCTTCTGCAATTTGTAAATGTCAC 780
|||||
OY 1385 G-CGCGTTTAAATCCAGAGG 1404
|||||
Db 781 GCGCGTTTAAATCCAGAGG 801
|||||

RESULT 9
B0689588 872 bp mRNA linear EST 15-JUL-2002
LOCUS B0689588
DEFINITION AGENCOURT_8345928 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6250957

5', mRNA sequence.
B0689588
B0689588.1 GI:21814904
EST.
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 872)
NIH-MGC <http://mgc.nci.nih.gov/>.

AUTHORS
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

COMMENT
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LUCM2394 row: n column: 14
High quality sequence stop: 654.

FEATURES
source
1..872
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6250957"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 192 a 258 c 247 g 174 t
ORIGIN

Query Match 29.9%; Score 783.2; DB 14; Length 872;

Best Local Similarity 99.4%; Pred. No. 7.5e-161;

Matches 796; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 605 GCCTCTCTCTCTCATATATACAGCAGCCTGTGTGGCCCAACATCGAGACAGCGGAG 664
|||||
Db 1 GCCTCTCTCTCTCATATATACAGCAGCCTGTGTGGCCCAACATCGAGACAGCGGAG 60
|||||
OY 665 GAGCGCGGCTGACCTTCTGCGCAAGTTTATCCATGTCTGTGATGACCCCAAGTCT 724
|||||
Db 61 GAGCGCGGCTGACCTTCTGCGCAAGTTTATCCATGTCTGTGATGACCCCAAGTCT 120
|||||
OY 725 GCGGCTGTGGCCACCTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 784
|||||
Db 121 GCGGCTGTGGCCACCTTGTCTATACAGAAAGTTCGACCGCTTACTGGGTACTGTGG 180
|||||
OY 785 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 844
|||||
Db 181 TGCCCCACAGCCTCTCTGGGAAGTTCAGAGGGCTCAAGACGCTCGAATCCTGTATGAG 240
|||||
OY 845 GAAATCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAG 904
|||||
Db 241 GAAATCGATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAG 300
|||||
OY 905 AAGAGGAGCTCGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCTTGA 964
|||||
Db 301 AAGAGGAGCTCGTATCGGTATACCCCAAGGACAGGACAGCAAGATCCCAAGTTCCTTGA 360
|||||
OY 965 CTGGGTGATTCACAGCTGACAGCGGCAAGATCGTCTGACCCAGAGAGAGAGAGCTG 1024
|||||
Db 361 CTGGGTGATTCACAGCTGACAGCGGCAAGATCGTCTGACCCAGAGAGAGAGAGAGCTG 420
|||||

QY	1025	GTGCAGCCCTTCAGCTGCCTGTCTCCGAAAGGTGAGTACATGCCAGGGCCGGGTGAGCC	1084
Db	421	GTGCAGCCCTTCAGCTGCCTGTCTCCGAAAGGTGAGTACATGCCAGGGCCGGGTGAGCC	480
QY	1085	CGGATGGCAAAATACGCGCTGGGSCATGTTCTCGAACCGGCCACAGTGGCTCCAGCTC	1144
Db	481	CGGATGGCAAAATACGCGCTGGGSCATGTTCTCGAACCGGCCACAGTGGCTCCAGCTC	540
QY	1145	GTCCCTTCCTCCCGCGGCCCTGTTCAATCCCGAGCACAGAGATGAGAGCAGCGGCTAGCC	1204
Db	541	GTCCCTTCCTCCCGCGGCCCTGTTCAATCCCGAGCACAGAGATGAGAGCAGCGGCTAGCC	600
QY	1205	TCTGCCAGAGCTGTGCCCAGGAATGCCAGCCGTATGAGTTAGGAGAGGTCAACCAAC	1264
Db	601	TCTGCCAGAGCTGTGCCCAGGAATGCCAGCCGTATGAGTTAGGAGAGGTCAACCAAC	660
QY	1265	GCTCGATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAAGAGGAGGACGAG	1324
Db	661	GCTCGATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAAGAGGAGGACGAG	720
QY	1325	CTCTGCTTTCTCCCGCGCCAAATGATGCAAGA-CGGGCTTCTGCCATTGTGTACAAAGTCAC	1383
Db	721	CTCTGCTTTCTCCCGCGCCAAATGATGCAAGAAGCCGGCTTCTGCCATTGTGTACAAAGTCAC	780
QY	1384	CGCGCTTTAAATCCCAAGG 1404	
Db	781	CGCGCTTTTAATCCCAAG 801	

```

RESULT 10
BM461814
LOCUS      BM461814               866 bp    mRNA    linear    EST 05-FEB-2002
DEFINITION AGNCOURT_6418408 NIH_MGC_71 Homo sapiens CDNA IMAGE:553867
            5', mRNA sequence.
ACCESSION  BM461814
VERSION    BM461814.1  GI:18510854
KEYWORDS   EST.
SOURCE      human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
            1 (bases 1 to 866)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1AM12219  row: 9  column: 20
            High quality sequence stop: 627.
FEATURES
            source
            Location/Qualifiers
            1..866

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:553867"
/clone_1ib="NIH_MGC_71"
/rissue_type="leiomyosarcoma"
/lab_host="DH10b (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 2.1 kb."
BASE COUNT      188 a      269 c      245 g      164 t
ORIGIN

Query Match      29.8%      Score 779.8;  DB 13;  Length 866;
Best Local Similarity  98.8%      Pired. No. 4,2e-160;

```

	Matches	81% Conservative	0 Mismatches	7 Indels	3 Gaps
QY	128	GGCAGCGCGAAGTACTGGGGCTCATTTGTCAACAAGCGCCCGCCACGACTTCCAGTTTGTG			187
Db	11	GGCAGCGCGAAGTACTGGGGCTCATTTGTCAACAAGCGCCCGCCACGACTTCCAGTTTGTG			70
QY	188	CAGAAGAAGATGATGTCGTGGGGCCCACTCCACCGCCCTCTACTACCTGGAAATGCATAT			247
Db	71	CAGAAGAAGATGATGTCGTGGGGCCCACTCCACCGCCCTCTACTACCTGGAAATGCATAT			130
QY	248	GGCAGCGCGAAGAACCTCCCTCTCTACTCTGAGATTTCCAAAGAAAGTCCGGAAAGAGCT			307
Db	131	GGCAGCGCGAAGAACCTCCCTCTCTACTCTGAGATTTCCAAAGAAAGTCCGGAAAGAGCT			190
QY	308	CTGCGTCCCTCTGCTGGAAACATAGTGGATATTTCCAGGCGCAGCCCGCCACCATGGG			367
Db	191	CTGCGTCCCTCTGCTGGAAACATAGTGGATATTTCCAGGCGCAGCCCGCCACCATGGG			250
QY	368	GTCCTACTCTCGGAGAGAGAGACTGCTGAGGAGGCGGAACGCTGGGGGTCTTCGGCATC			427
Db	251	GTCCTACTCTCGGAGAGAGAGACTGCTGAGGAGGCGGAACGCTGGGGGTCTTCGGCATC			310
QY	428	ACCTCTACGACTTCACACAGCGAGAGTGGCTCTTCTCTTCAGGCGCAGCAACAGCCTC			487
Db	311	ACCTCTACGACTTCACACAGCGAGAGTGGCTCTTCTCTTCAGGCGCAGCAACAGCCTC			370
QY	488	TTTCACTGTGCGAGCGGGGCGAAGAGCGGCTTCATGGTGTCCCTATGAAACCGCTGGAA			547
Db	371	TTTCACTGTGCGAGCGGGGCGAAGAGCGGCTTCATGGTGTCCCTATGAAACCGCTGGAA			430
QY	548	ATCAAGAACCAGTGTCTGAGGGCCCGGAGTGGAGCCCAAAATCTGGCTTCGCCAGCCTGCC			607
Db	431	ATCAAGAACCAGTGTCTGAGGGCCCGGAGTGGAGCCCAAAATCTGGCTTCGCCAGCCTGCC			490
QY	608	TTCTCTCTCTTCAATCAATTAACAGAGACTGTGGGTGGGCGCAACATTCGAAACAGGAGAG			667
Db	491	TTCTCTCTCTTCAATCAATTAACAGAGACTGTGGGTGGGCGCAACATTCGAAACAGGAGAG			550
QY	668	CGGGGGGTGACCTTCTGCGACCAAGGTTTATCCAAATGCTCGATGAGACCCCAAGTCTCGG			727
Db	551	CGGGGGGTGACCTTCTGCGACCAAGGTTTATCCAAATGCTCGATGAGACCCCAAGTCTCGG			610
QY	728	GGTGTGGCCACCTTCGTGCATACAGAGAGATTTCAGCGGCTTCACTGGGTACTGTTGTGC			787
Db	611	GGTGTGGCCACCTTCGTGCATACAGAGAGATTTCAGCGGCTTCACTGGGTACTGTTGTGC			670
QY	788	CCCAAGGCTCTCTGGGAAGTTTCAGAGGGGCGTCAAGAGAGTGCAGAAATCTGTATGAGAA			847
Db	671	CCCAAGGCTCTCTGGGAAGGTCTCAGAGGGGCTCAAGAGAGTGCAGAAATCTGTATGAGAA			730
QY	848	GTCGATGAGT -CCGAGGTGAGAGTCAATTCACGTCGCCCTCTCTGGCTAGAAAG -AAAAG			905
Db	731	GTCGATGAGTCCCGAGGTGAGAGTCAATTCACGTCGCCCTCTCTGGCTAGAAAGAAAGGA			790
QY	906	AGAGGGACTCTATTCGGTA -CCCCAGAGCAGGCAAGCAAGAAATCCCA 951			
Db	791	AGAGGGACTCTATTCGGTATCCCCAGAAACAGGCAAGCAAGAAATCCCA 837			

RESULT	11
BQ949519	
LOCUS	
DEFINITION	BQ949519 910 bp mRNA linear EST Z1-ANUG-2002
ACCESSION	AGENCOURT_8794758 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374913
VERSION	5 , mRNA sequence.
KEYWORDS	BQ949519 BQ949519.1 GI:22364997
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 910)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

```

Plate: LLCM2554 row: k column: 10
High quality sequence stop: 668.
FEATURES
Location/Qualifiers
source
1..910

```

Db	577	AAGACCCAGTGTCTAGGGGCCCCGGATGTGATGCCCAAAATCTGCGCTGGCAGCCCTTCC	636
Qy	611	TTTCTCTTCATCAATTAAGAGCGACCTGTGGGTGGGCAACATCGAGACAGGGAGAGCGG	670
Db	637	TTTCTCTTCATCAATTAAGAGCGACCTGTGGGTGGGCAACATCGAGACAGGGAGAGCGG	696
Qy	671	CGGCTGACCTTGTGGCCACCAAGTTTATCCAAATGTCTCGATGATGACCCCAAGTCTGC	729
Db	697	CGGCTGACCTTGTGCA -CAAGTTTATCCAAATGTCTCGATGATGACCCCAAGTCTGC	755
Qy	730	TGTGGCACTTCGTATACAGAGAGAGTGC -CCGCTTCACTGGTACTGTGTGCTCC	788
Db	756	TGTGGCACTTCGTATACAGAGAGAGTGCACCCGCTTCACTGGTACTGTGTGCTCC	815
Qy	789	CCACG -CCTCTGGGAAGTTCAGAGGGCTCAAGACGTG -CGATCTGTATGAGA	846
Db	816	CCACGACCTCTCGGGGAGTTCAGAGGGCTCAGACGCTGCGAATCTGTATGTAGA	875
Qy	847	AGTCGATGATCC	859
Db	876	AGTCGATGATCC	888

RESULT 12	
BQ642814	
LOCUS	902 bp mRNA linear EST-15-JUL-2002
DEFINITION	AGENCOURT_8285906 NIH_MGC__99 Homo sapiens CDNA clone IMAGE:6292546
ACCESSION	BQ642814 BQ642814.1 GI:21766986
VERSION	EST.
KEYWORDS	
SOURCE	human.

[illegible]

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 1051)	Enkaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
1	NH-MGC	http://mgc.nci.nih.gov/.		
1	National Institutes of Health.	Mammalian Gene Collection (MGC)		
1	Unpublished (1999)	Contact: Robert Strausberg, Ph.D.		
1	Email: cgaphs-remail.nih.gov			
1	Tissue Procurement: DCTD/DRP			
1	cDNA Library Preparation: Rubin Laboratory			
1	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)			
1	DNA Sequencing by: Agencourt Bioscience Corporation			
1	clone distribution: MGC clone distribution information can be			
1	found through the I.M.A.G.E. Consortium/LNL at:			
1	http://image.lnl.gov			
1	Plate: LNCM2007	row: 0 column: 06		
1	High quality sequence stop: 677.			
1	Location/Qualifiers			
1	1..1051			
1	/organism="Homo sapiens"			
1	/db_xref="taxon:9606"			
1	/clone="IMAGE:5482205"			
1	/issue_lib="NIH_MGC_41"			
1	/tissue_type="amelanotic melanoma, cell line"			
1	/lab_host="DH10B (phage-resistant)"			
1	/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:			
1	EcotRI; cDNA made by oligo-dT priming. Directionally cloned			
1	into EcotRI/XhoI sites using the following 5' adaptor:			
1	GGCCAGG(G). Library constructed by Ling Hong in the			
1	Laboratory of Gerald M. Rubin (University of California,			
1	Berkeley) using ZAP-cDNA Synthesis Kit (Stratagene) and			
1	Superscript II RT (Life Technologies). Note: this is a			
1	NIH_MGC Library."			
BASE COUNT	233 a	329 c	300 g	189 t
ORIGIN				
Query Match	29.3%	Score 767.4:	DB 14:	Length 1051:
Best Local Similarity	96.0%:	Pred. No. 2.3e-157:		
Matches 853:	Conservative 0:	Mismatches 26:	Indels 10:	Gaps 6:
QY	349	GGCCAGCCCCACACATGGGGTCTACTCTTCGGGAGAGAGAGCTGTGAGGAGCGGAACG	408	
Db	1	GGCCAGCCCCACACATGGGGTCTACTCTTCGGGAGAGAGAGCTGTGAGGAGCGGAACG	60	
QY	409	CGTGGGGTCTTCGGGATACCTCTTCAGACTTCCACAGCGAGTGGCTTCTCTT	468	
Db	61	CGTGGGGTCTTCGGGATACCTCTTCAGACTTCCACAGCGAGTGGCTTCTCTT	120	
QY	469	CCAGGCGAGCAACAGCTCTTCCACTGTTCGAGCGGCGCAAGAACGGCTTATGCTGTC	528	
Db	121	CCAGGCGAGCAACAGCTCTTCCACTGTTCGAGCGGCGCAAGAACGGCTTATGCTGTC	180	
QY	529	CCCATGGAACCGCGTGAATCAAGACCCAGTGGCTCAGGGCCCGGATGGAGCCCAAAAT	588	
Db	181	CCCATGGAACCGCGTGAATCAAGACCCAGTGGCTCAGGGCCCGGATGGAGCCCAAAAT	240	
QY	589	CTGGCTCCGACACCTCGCTCTTCTCCCTTATCAATTAACAGCGACTGTGGGTGGCAA	648	
Db	241	CTGGCTCCGACACCTCGCTCTTCTCCCTTATCAATTAACAGCGACTGTGGGTGGCAA	300	
QY	649	CATGAGACAGGCGAGAGCGGCGGCTGACCTTCTGCACCAAGGTTTATCCATATGCTCT	708	
Db	301	CATGAGACAGGCGAGAGCGGCGGCTGACCTTCTGCACCAAGGTTTATCCATATGCTCT	360	
QY	709	GGATGACCCCAAGTCTGGGGGTGTGGCCACCTTGCTCATACAGAAAGATTGACCGCTT	768	
Db	361	GGATGACCCCAAGTCTGGGGGTGTGGCCACCTTGCTCATACAGAAAGATTGACCGCTT	420	
QY	769	CACGTGGATGAGTGTGGTCCCGCACAGCTCTCGGGGAAGGTTTCAGAGGGCTCAAGAGCT	828	
Db	421	CACGTGGATGAGTGTGGTCCCGCACAGCTCTCGGGGAAGGTTTCAGAGGGCTCAAGAGCT	480	
QY	829	GCGAATCCCTGATGAGGAAGTGAATGAGTCCGAGGTGAGGTCATTTCAAGTCCCTTCC	888	

|||||
Db 481 GCGAATCCTATGAGAGAAAGTCGATGAGTCCGAGGTGAGAGTCAATACGTCCTCC 540
Qy 889 TGCCTAGAAAAGAAAGAGAGAGCTGTATCGTATGCCCGAGAGAGCAGCAAGATCC 948
Db 541 TGCCCTAGAAAAGAAAGAGAGAGCTGTATCGTATGCCCGAGAGAGCAGCAAGATCC 600
Qy 949 CAGATTGCTCTTGAACAGGCTGATGTTCCAGCTGACAGCCAGGCAAGATGCTGTGAC 1008
Db 601 CAGATTGCTCTTGAACAGGCTGATGTTCCAGCTGACAGCCAGGCAAGATGCTGTGAC 660
Qy 1009 CCAGAGAGAGAGCTGTGTCAGCCCTTCAGCTGCTGTTCGGAAGTGAATACATGCG 1068
Db 661 CCAGAGAGAGAGCTGTGTCAGCCCTTCAGCTGCTGTTCGGAAGTGAATACATGCGCA 720
Qy 1069 CAGGCGCGGTG--ACCCGGGATGCAATAGC-CCTGGGCATGTTCTGACCGG- 1123
Db 721 GGGGGGCGGCTGGAAACCCCGGAATGGCAATAGCCCTGGGCGCATGTTCTGGACCGG 780
Qy 1124 CCCAGCAGTGGCTCCAGCTGCTCTCTCCGCC-GGCCCTGTTCAATCCGAGAC--A 1180
Db 781 CCCAGCAGTGGCTCCAGCTGCTCTCTCCGCCCGGCGCTGTTCAATCCGAGACCCCA 840
Qy 1181 GAGATGAGAGAGAGAGCT--AGCTTCCAGAGAGTGTCCCGAGAA 1227
Db 841 AGAATGAGAGAGAGAGCTTACCTCTGCGAGAGACTGGCCCAAGAA 889

RESULT 14
BI223668 787 bp mRNA linear EST 11-Jul-2001
LOCUS 602943581.F1 NIH_MGC_12 Homo sapiens cdna clone IMAGE:5106357 5',
DEFINITION mRNA sequence.
ACCESSION BI223668
VERSION BI223668.1 GI:14677112
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 787)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNLT Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgarbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L14M1257 row: b column: 22
High quality sequence stop: 787.

FEATURES
source location/Qualifiers
1..787

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5106357"
/clone_1id="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT 164 a 263 c 220 g 140 t
ORIGIN

Query Match 29.0%; Score 757.8; DB 13; Length 787;
Best Local Similarity 99.0%; Pred. No. 2,6e-155;
Matches 773; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 11 ATGGCCACACCGGGACCCCAACGGCCGAGCAGGCGCAGCAGCCGACAGATGACCG 70
Db 7 ATGGCCACACCGGGACCCCAACGGCCGAGCAGGCGCAGCAGCCGACAGATGACCG 66
Qy 71 GCCCGCCCTTCCAGGTGACAGAACACTGCTGGAGACGGGCTCCGAGCATATCCAGCG 130
Db 67 GCCCGCCCTTCCAGGTGACAGAACACTGCTGGAGACGGGCTCCGAGCATATCCAGCG 126
Qy 131 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCACAGCTTCCAGTTGTGAG 190
Db 127 AGCCGCAAGTACTGGGCTCATTTGTCAACAAGGCGCCACAGCTTCCAGTTGTGAG 186
Qy 191 AAGAGGATGAGTCTGGGCCCCCATCCACCGGCTCTACTACTCTGGGAATCCATATGCG 250
Db 187 AAGAGGATGAGTCTGGGCCCCCATCCACCGGCTCTACTACTCTGGGAATCCATATGCG 246
Qy 251 AGCCGAGAACTCCCTCTACTCTGAGATTCCCAAGAGGTCGGAAAGAGAGCTCTG 310
Db 247 AGCCGAGAACTCCCTCTACTCTGAGATTCCCAAGAGGTCGGAAAGAGAGCTCTG 306
Qy 311 CTGCTCTGTCTGTGGAAGCAGATCTGTGATCATTTCCAGGCCACGCCACCATGGGGTC 370
Db 307 CTGCTCTGTCTGTGGAAGCAGATCTGTGATCATTTCCAGGCCACGCCACCATGGGGTC 366
Qy 371 TACTCTGGGAGAGAGAGCTCTGAGGAGCGGAAAGCCCTGGGGGTCTTGGCATCAC 430
Db 367 TACTCTGGGAGAGAGAGCTCTGAGGAGCGGAAAGCCCTGGGGGTCTTGGCATCAC 426
Qy 431 TCCTACGACTTCCACAGGAGAGTGGCTCTTCCCTTCAGGCCAGCAACAGCTCTTC 490
Db 427 TCCTACGACTTCCACAGGAGAGTGGCTCTTCCCTTCAGGCCAGCAACAGCTCTTC 486
Qy 491 CACTGTGCGAGCGGCGGCAAGACGGCTTCATGTGTCCCTATGAAACCGCTGAAATC 550
Db 487 CACTGTGCGAGCGGCGGCAAGACGGCTTCATGTGTCCCTATGAAACCGCTGAAATC 546
Qy 551 AAGACCCAGTCTCAGGGCCCCCGATGAGACCCCAAAATCTGCTCCGCGACCTGCTTC 610
Db 547 AAGACCCAGTCTCAGGGCCCCCGATGAGACCCCAAAATCTGCTCCGCGACCTGCTTC 606
Qy 611 TTCTCTCTCATCAATAACAGGAGCTGTGGTGCGCAACATGAGAGAGCGGAGAGCGG 670
Db 607 TTCTCTCTCATCAATAACAGGAGCTGTGGTGCGCAACATGAGAGAGCGGAGAGCGG 666
Qy 671 CGCTGACCTTCTGCGACCAAGTTTATCAATGTCTGATGACCCCAAGTCTGCGGCT 730
Db 667 CGCTGACCTTCTGCGACCAAG--TTATCAATGTCTGATGACCCCAAGTCTGCGGCT 725
Qy 731 GTGGCCACTTCGTCATACAGAGAGTTGACACCGCTTCACTGGGTACTGTGTCGCC 790
Db 726 GTGGCCACTTCGTCATACAGAGAGTTGACACCGCTTCACTGGGTACTGTGTCGCC 785
Qy 791 A 791
Db 786 A 786

RESULT 15
B0685669 938 bp mRNA linear EST 15-Jul-2002
LOCUS B0685669
DEFINITION AGNCOURT_8176462 NIH_MGC_110 Homo sapiens cdna clone IMAGE:6252323
5', mRNA sequence.
ACCESSION B0685669
VERSION B0685669.1 GI:21810985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 938)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
plate: L1CM2398 row: g column: 12
High quality sequence stop: 646.

BASE COUNT	209 a	278 c	262 g	189 t
ORIGIN				

Query Match	28.7%	Score 750:	DB 14:	Length 938:
Best Local Similarity	97.9%	Pred. No. 1.4e-153:		
Matches 781; Conservative	0;	Mismatches 15;		Indels 2; Gaps 2

QY	605	GCCTTCCTCCTTCATCAATAAACAGGACCTGTGGGTGGCAACATCGAGACGGGAG	664
Db	1	GCCTTCCTTCCTTCATCAATAAACAGGACCTGTGGGTGGCAACATCGAGACGGGAG	60
QY	665	GAGCGGGGGCTGACCTTCTGGCACCAAGTTTATCAATGTCTGTGATGACCCCAAGCT	724
Db	61	GAGCGGGGGCTGACCTTCTGGCACCAAGTTTATCAATGTCTGTGATGACCCCAAGCT	120
QY	725	GGGGGTGGGGCACCTTCGTATCAAGGAAGAGTTGACCGGCTCACTGGGTACTGGTG	784
Db	121	GGGGGTGGGGCACCTTCGTATCAAGGAAGAGTTGACCGGCTCACTGGGTACTGGTG	180
QY	785	TGCCCCACAGCCCTCCTGGGAAGTTTCAGAGGGCTTCAAGACGCTGCANATCTGTATGAG	844
Db	181	TGCCCCACAGCCCTCCTGGGAAGTTTCAGAGGGCTTCAAGACGCTGCANATCTGTATGAG	240
QY	845	GAATTCATAGTCCGAGAGTGGAGAGTATTCACGTCCCTCTCTGTGGCTAGAAAGAAAG	904
Db	241	GAATTCATAGTCCGAGAGTGGAGAGTATTCACGTCCCTCTCTGTGGCTAGAAAGAAAG	300
QY	905	AAGACGAGACTCGTATCCAGTACCCAGAGACAGCGCAAGATCCCAAGATTGCTTGA	964
Db	301	AAGACGAGACTCGTATCCAGTACCCAGAGACAGCGCAAGATCCCAAGATTGCTTGA	360
QY	965	CTGGCTGAGTTCAGACTGACAGCGCAGGGCAAGATGCTTCGACCCAGGAAGAGACTG	1022
Db	361	CTGGCTGAGTTCAGACTGACAGCGCAGGGCAAGATGCTTCGACCCAGGAAGAGACTG	420
QY	1025	GTCGAGCCCTTCACTGTGGCTGTTCCGAAGTGAAGTACATACGCCA-GGGCCGGGTGAC	108
Db	421	GTCGAGCCCTTCACTGTGGCTGTTCCGAAGTGAAGTACATACGCCAAGGGCCGGGTGAC	480
QY	1084	CCGGATGAGCAAAATACGCTGTGGGCAATGTTCTTGACCGGCCCCAGCAGTGGCTCAGCT	1144
Db	481	CCGGATGAGCAAAATACGCTGTGGGCAATGTTCTTGACCGGCCCCAGCAGTGGCTCAGCT	540
QY	1144	CGTCTCTCTCCCCCGGCTCTTTCATCCGAGACAGAGAAATGAGGACGCGCTAGC	120

Db	541	CGTCTCTCTCCCCCGGCGCTGTTCATCCGAGCACAGAGAAATGAGGAGCAGCGGCTAGC	600
Qy	1204	CTCTGCCAGAGCTGTCTCCCAAGGAATGTCCAGCCGTTATGTGTGTACGAGGAGTACC	1265
Db	601	CTCTGCCAGAGCTGTCTCCCAAGGAATGTCCAGCCGTTATGTGTGTAGAGGAGGTTAC	660
Qy	1264	CGTCTGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCAGA-GGAGAGAG	1322
Db	661	CGTCTGATCAATGTTCATGACATCTTCTATCCCTTCCCCCAATCAAGGGAGAGGAA	720
Qy	1323	AGCTCTGCTTTCTCCGCGCAATGAATGCAAGACCGGCTTGTGCCATTGTTCAAATG	1388
Db	721	AGCTCTGCTTTCTCCGCGCAATGAATGCAAGACCGGCTTGTGCCATTGTGCAAAATG	780
Qy	1383	CCGCGGTTTAAATCC 1400	
Db	781	CCGCGGTTTAAATCC 798	

Search completed: December 12, 2002, 10:36:46
Job time : 2402 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 08:32:37 : Search time 74 Seconds
(without alignments)
10845.582 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 2617

Sequence: 1 caagcttaccatgcccaccca.....tctgagcgccgcggtatccg 2617

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCPUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152.8	5.8	823	4	US-09-280-116-171
2	127	4.9	612	4	US-09-392-184-31
3	87.6	3.3	5496	4	US-09-462-284-1
4	68.2	2.6	543	4	US-09-221-017B-253
5	55.6	2.1	2924	5	PCT-US93-07923-1
6	46.6	1.8	2815	1	US-08-230-91A-1
7	46.6	1.8	2815	1	US-08-619-280A-1
8	46.6	1.8	2815	2	US-08-940-391-1
9	45.4	1.7	1575	3	US-08-858-876A-1
10	45.4	1.7	1575	4	US-09-472-880-1
11	45.4	1.7	8438	1	US-07-945-283-1
12	45	1.7	1342	3	US-08-832-399-1
13	45	1.7	1342	4	US-09-372-498-1
14	44.4	1.7	657	4	US-09-221-017B-646
15	44.4	1.7	3624	1	US-07-951-715A-6
16	44.4	1.7	3624	2	US-08-459-448A-6
17	44.4	1.7	3624	3	US-08-459-595A-6
18	44.4	1.7	3624	3	US-08-459-504B-6
19	44.4	1.7	3624	3	US-08-459-444-6
20	44.4	1.7	3624	3	US-09-053-549-7
21	44.4	1.7	3624	4	US-09-547-422-6
22	44.4	1.7	8854	3	US-09-053-549-1
23	43.8	1.7	2214	3	US-08-864-038A-1
24	43.8	1.7	3331	3	US-08-864-038A-2
25	43.8	1.7	3331	3	US-08-864-038A-4
26	43.6	1.7	43280	2	US-08-804-227C-1
27	43.2	1.7	962	4	US-08-765-907A-16

28	43.2	1.7	1052	2	US-08-403-852D-10	Sequence 10, Appl
29	43.2	1.7	1052	3	US-08-510-646B-10	Sequence 10, Appl
30	43.2	1.7	1052	4	US-09-231-818-10	Sequence 10, Appl
31	42.2	1.6	1835	4	US-09-417-704-2	Sequence 2, Appl1
32	42	1.6	7898	4	US-08-984-709A-49	Sequence 49, Appl
33	42	1.6	49272	1	US-08-614-770A-1	Sequence 1, Appl1
34	41.8	1.6	2849	3	US-08-809-286B-1	Sequence 1, Appl1
35	41.6	1.6	926	4	US-09-500-569-19	Sequence 19, Appl
36	41.4	1.6	8051	2	US-08-576-626A-2	Sequence 2, Appl1
37	41.2	1.6	1829	1	US-08-035-558-2	Sequence 2, Appl1
38	41.2	1.6	1829	2	US-08-682-847-3	Sequence 2, Appl1
39	41	1.6	1302	4	US-09-372-422A-27	Sequence 27, Appl1
40	41	1.6	4960	4	US-09-907-843-3	Sequence 3, Appl1
41	40.8	1.6	1554	1	US-08-469-486-1	Sequence 1, Appl1
42	40.8	1.6	1554	2	US-08-469-658-1	Sequence 1, Appl1
43	40.8	1.6	2481	1	US-08-324-243-35	Sequence 35, Appl
44	40.8	1.6	2481	1	US-08-532-390-35	Sequence 35, Appl
45	40.8	1.6	2481	3	US-08-717-294-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-09-280-116-171
: Sequence 171, Application US/09280116A
: Patent No. 6331427
: GENERAL INFORMATION:
: APPLICANT: Robison, Keith E.
: TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
: FILE REFERENCE: 5800-24, 035800/176965
: CURRENT APPLICATION NUMBER: US/09/280, 116A
: CURRENT FILING DATE: 1999-03-26
: NUMBER OF SEQ ID NOS: 268
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 171
: LENGTH: 823
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-171

QY	2103	GCCAGGTGAGATGAGAGACGAGTGG-AGGCGCTGAGTTCGTGGCCGAGAGATGAGC	2161	Sequence 171, Appl
DB	1	CTCAATATGAATATGACGATGAGTGGAGACATTCATATCTAGCTTCGATATGAT	60	Sequence 31, Appl
QY	2162	TTTCATGACCTGAGCCGAGTTCATCCATGCGTGTCTTACGAGGGCTTCCTGCTC	2221	Sequence 253, App
DB	61	TTTCATGACTTAGATGCGTGGGATCCAGCGTGTCTTATGAGATATCTCTCCCTG	120	Sequence 1, Appl1
QY	2222	ATGGGGCTTAATCCAGACCCAGGTGTTCAAGTGGCATTCGGGGTCCCGGTGACC	2281	Sequence 1, Appl1
DB	121	ATGGCATTAATGACGAGGTGAGATATCTTCAGGGTGTATTTGGTCCCGCATCT	180	Sequence 1, Appl1
QY	2282	GTCGTGATGCGCTGACACAGGATGACACTGAGCGCTTACGAGCGTCCGAGAACAC	2341	Sequence 1, Appl1
DB	181	CTGTGATCTTCTTATGATACAGGATACGAGACCTTATATGGGTCCACCTGACAGAT	240	Sequence 6, Appl1
QY	2342	CAGCAGCGCTATGAGCGGTTCCGTGCGCTGACGAGTGAAGACCTGCCAATGAGCC	2401	Sequence 6, Appl1
DB	241	GACAGGCGCTTATCTTATGATGATCTGTGGCATGACGAAGAAGTTCCCTGACCA	300	Sequence 6, Appl1
QY	2402	AACGCTGCTTATCTTACGAGGCTCTCTGACGAGAAACGTCGACTTTTCCACCAAC	2461	Sequence 6, Appl1
DB	301	AATGCTTACTGCTCTTACATGATGTTCTCTGATGAGATGTCATTGTGACATACAGT	360	Sequence 1, Appl1
QY	2462	TTCTGCTGCTCCCACTGATCCGAGGAGAAACCTTACGAGTTCAG	2509	Sequence 16, Appl

Db 361 ATATACTGAGTTTTTTAGTAGAGGCGTGAAGACGATGATTTTACAG 408

RESULT 2

US-09-392-184-31/c

; Sequence 31, Application US/09392184

; Patent No. 6395889

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: PROTEASE HOMOLOGS

; FILE REFERENCE: 5800-55

; CURRENT APPLICATION NUMBER: US/09/392,184

; CURRENT FILING DATE: 1999-09-09

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 31

; LENGTH: 612

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(612)

; OTHER INFORMATION: prolyl oligopeptidase)

; NAME/KEY: misc_feature

; LOCATION: (1)...(612)

; OTHER INFORMATION: n = A,T,C or G

US-09-392-184-31

Query Match

Best Local Similarity 4.9%; Score 127; DB 4; Length 612;

Best Local Similarity 56.0%; Pred. No. 1.6e-19;

Matches 228; Conservative 0; Mismatches 178; Indels 1; Gaps 1;

QY 2157 ATGGCTCATCGACCGCGAGTGGCCATCGCTGCTGCGGCGCTTCTCT 2216

DB 472 ATATATTCATTGACNTAAGATCGTGGCAATCCGCGCTGCTATGAGGATACCTCT 413

QY 2217 CGCTCATGGGGCTAATTCACACAGCCCGAGTGTTCAGGTGGCCATCGCGGCGCCG 2276

DB 412 CCTGTATGGGANTAAATTCAGAGTCAAGATATCTMTCAAGGTT-CTATTTCTGGGGCCCGCAG 354

QY 2277 TCACCGTCTGGATGGCTTACGACAGAGGTACACTGAGCCCTCAGTGGAGCTCCCTGAGA 2336

DB 353 TCACCTGTGGATCTTCTTAAGAAACAGAAACGCGGACGCTTANATGGGTCAACCCCTNNCC 294

QY 2337 ACACAGCAGCAGCGCTAAGAGCGGTTCCGTGGCCCTGCAGCTGAGAGAGCTGCCAATG 2396

DB 293 AGAATGACAGGCGCTATTTACTTAGATCTGTGCCATTCACANAGGAAAGTTCCCTCTG 234

QY 2397 AGCCCAACGCGTTCCTTATCTCCACGCGCTTCTGAGAGAAAGTGACATTTTCCACA 2456

DB 233 NACCCATNNGTTCCTCTTANATGGTTCCTGGATTTGGAATTTCCANTTTTCCACANA 174

QY 2457 CAACATCTCTCGTCTCCCAACTGATCCGAGCAGGAGAACTTCCAGCTCCAATCTACC 2516

DB 173 CCGANNANATCTAGATTTTGTAGAGGCGTGGGAAACCAATGTGATTTCAACCAAC 114

QY 2517 CCAACGAGACAGATATTCGCTGCCGAGTGGGCGGAGCACTAT 2563

DB 113 CTCNGAGAGAACCAACANNAGAGTCTGATCGGAGGACAAATAT 67

RESULT 3

US-09-462-284-1

; Sequence 1, Application US/09462284

; Patent No. 6309868

; GENERAL INFORMATION:

; APPLICANT: Nestec S.A.

; APPLICANT: Monod, Michel

; APPLICANT: Donnas, Agnes

; APPLICANT: Affolter, Michael

; APPLICANT: Van Den Broek, Peter

; TITLE OF INVENTION: CLONING OF THE

; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM

; TITLE OF INVENTION: ASPERGILLUS ORYZAE

; FILE REFERENCE: 8265-298

; CURRENT APPLICATION NUMBER: US/09/462,284

; CURRENT FILING DATE: 2000-01-03

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 5496

; TYPE: DNA

; ORGANISM: Fungus

US-09-462-284-1

Query Match

Best Local Similarity 3.3%; Score 87.6; DB 4; Length 5496;

Best Local Similarity 49.1%; Pred. No. 1.6e-10;

Matches 321; Conservative 0; Mismatches 324; Indels 9; Gaps 3;

QY 1840 CTTTCATTTCCACAGCGCTCGATGTGCGGCTCTACGCGCATGATCTACAGCCCGCACGC 1899

DB 3412 CTTTCGAGCTTCCCTCCCTCCCGAGAAACCTCAATGTATGACGCGCTTACCCCGCG 3471

QY 1900 CTTTCGAGCGGAGAGAGACCCACCGCTCTTTGTATGAGAGCCCGCCAGGTGCA 1959

DB 3472 GTTCTCCCGGATAGAGAGATGATCTTACCCCATACCCCTACGCGCGCCAGCGCCCA 3531

QY 1960 GTTGTGATATATCTCTTCAAGCATCATGACTTTCGCGCTCAACACACTGCGCTCCT 2019

DB 3532 AGAATGACCAAGATGAG 3591

QY 2020 GGGCTACGCGGCTGTGTATGAGAGCA--GGGCTCTCTGTACAGAGAGGCTTGGTT 2076

DB 3592 ACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3651

QY 2077 CGAAGGAGGCGCTGAAAGCAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2136

DB 3652 CGGCTCGCGTACAGCGGCTGAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 3711

QY 2137 GCATTCGTGGCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2196

DB 3712 GCAGAGGCGGCGCAACATGCC---CTGATGATGATGATGATGATGATGATGATGATG 3768

QY 2197 GTTCATGAGGCGGCT 2256

DB 3769 GACTTTCGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3828

QY 2257 GGCATTCGCGGAGGCGCGGCTGACGCTGATGATGATGATGATGATGATGATGATGATG 2316

DB 3829 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3888

QY 2317 CTACATGAGAGCTCTGAG 2376

DB 3889 CTACATGAGAGAGCTCTGAG 3945

QY 2377 CTTGAGAGAGCTCCCAAG 2436

DB 3946 GACTTCGAGGCTTCAAG 4005

QY 2437 AAAGTGCAGCTTTTCCACAGCAAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2490

DB 4006 TAACTGCATTTCCAGAACTGCGCTGCGTGTGATGATGATGATGATGATGATGATG 4059

RESULT 4

US-09-221-017B-253

; Sequence 253, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: Ross, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

Db 1910 TACGTAACCTCAATGCTCTCGGATCAGGAAGTGGCGTTCAGTGTGGAATAGCCGTG 1969
Oy 2270 GCCCGGTCACCGTGTGATGGGCTTCAGACACAGGAGTACAGGCTACATGAGACGTC 2329
Db 1970 GCGCCTGTATCCCGGTGGAGAGTACTATGACTAGTACACAGAACGTTACATGGCTCTC 2029
Oy 2330 CC 2331
Db 2030 CC 2031

RESULT 6
US-08-230-491A-1
; Sequence 1, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Retlig, Wolfgang J.; Scanlan, Matthew J.;
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
; NUMBER OF INVENTIONS: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FELFE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-230-491A-1

Query Match 1.8%; Score 46.6; DB 1; Length 2815;
Best Local Similarity 54.3%; Pred. No. 0.18;
Matches 94; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Oy 2159 GGCTTCATGACCTGACGCGAGTGGCATCCATGCGTGTCTCCAGGGGGCTTCCTCG 2218
Db 2039 GGTTCATTGATGAGAAAAAGAAATAGCCATATGCGGCTGTCTCATGAGGATACGTTCA 2098
Oy 2219 CTCATGGGCTATATCCACAAGCCCAAGTGTTCAGAGTGGCCATCGGGGTGCCCGGTC 2278
Db 2099 TCACGTGGCCCTTGCACTGTGGAAGTGTCTTTCAATGTGATATACAGTGGGTCCAGTC 2158
Oy 2279 ACCGCTGTGAGGCGCTACGACACAGGGTACACTGAGCGCTACATGAGAGCTCC 2331
Db 2159 TCACACTGGGAAATATATACGCTGTGCTACACAGAGAGATTCATGGGTCTCC 2211

RESULT 7
US-08-619-280A-1
; Sequence 1, Application US/08619280A

; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felife & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-619-280A-1

Query Match 1.8%; Score 46.6; DB 1; Length 2815;
Best Local Similarity 54.3%; Pred. No. 0.18;
Matches 94; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Oy 2159 GGCTTCATGACCTGACGCGAGTGGCATCCATGCGTGTCTCCAGGGGGCTTCCTCG 2218
Db 2039 GGTTCATTGATGAGAAAAAGAAATAGCCATATGCGGCTGTCTCATGAGGATACGTTCA 2098
Oy 2219 CTCATGGGCTATATCCACAAGCCCAAGTGTTCAGAGTGGCCATCGGGGTGCCCGGTC 2278
Db 2099 TCACGTGGCCCTTGCACTGTGGAAGTGTCTTTCAATGTGATATACAGTGGGTCCAGTC 2158
Oy 2279 ACCGCTGTGAGGCGCTACGACACAGGGTACACTGAGCGCTACATGAGAGCTCC 2331
Db 2159 TCACACTGGGAAATATATACGCTGTGCTACACAGAGAGATTCATGGGTCTCC 2211

RESULT 8
US-08-940-391-1
; Sequence 1, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felife & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-OCT-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-940-391-1

Query Match 1.8%; Score 46.6; DB 2; Length 2815;
Best Local Similarity 54.3%; Pred. No. 0.18; Indels 0; Gaps 0;
Matches 94; Conservative 0; Mismatches 79;

QY 2159 GCGTCATGACCTGAGCCGAGTGGCCATGCTGCTACAGGGGCTTCCTCG 2218
DB 2039 GGTTCATGATGATAAAAGATAGCCATGATGGGCTGCTATGAGATAGCTTCA 2098
QY 2219 CTGATGGGCTATCCACAGACCCCGAGTGTTCAGGTGCGCATCGCGGCTGCC 2278
DB 2099 TCACGTGCGCTTCATGCTGAGACGTGCTTTCAATGTGGTATAGAGAGGCTCCAGTC 2158
QY 2279 ACCGTGTGATGGCTACGACACAGGCTACAGGCTGCTACATGACGCTCC 2331
DB 2159 TCCAGCTGGGATATTACGCGCTGTCTACACAGAGATTCATGGGCTCC 2211

RESULT 9
US-08-858-876A-1
Sequence 1, Application US/08858876A
Patent No. 6022856
GENERAL INFORMATION:
APPLICANT: Daniel CAPUT
APPLICANT: Pascale CHALON
APPLICANT: Pascual FERRARA
APPLICANT: VITA NATALIO
TITLE OF INVENTION: Type 2 Neurotensin Receptor
TITLE OF INVENTION: (hnt-R2)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh Street
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,876A
FILING DATE: 19-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 9723204
FILING DATE: 17-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,049
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 37..1266
US-08-858-876A-1

Query Match 1.7%; Score 45.4; DB 3; Length 1575;
Best Local Similarity 46.6%; Pred. No. 0.29;
Matches 145; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1506 GGCACGGCTCCAGATGTGGGTCAATGAGAGACCAAGCTGTACTTCCAGGGCACCA 1565
DB 119 GCGCTGGGGCCCAAGTGTCTTACCCGCTTACGACATCATCTGGGCGCCGCGG 178
QY 1566 AGGACGCGCGGCTGGAGACACACCTCTAGCTGAGTGTGACACTGTAGAGGGCGGAGATG 1625
DB 179 GGGCATGTGCTGTCTCCGTACAGTGTGTCTGAAGCGCGGCGGCGGCGGCGG 238
QY 1626 TAGGCTTACACAGCCCGGCTTCCCATAGCTGTCTCATGAGGACAGAACTTGACATGT 1685
DB 239 TGGCCACACAGTGTCTGCTGAGCTGGCGCTCGCGGCGCTGCTGCTGTGCTGCTG 298
QY 1686 TCGTACGCACTACAGAGGCTGAGACAGCGCGGCTGCTGCTGCTGCTGCTGCTG 1745
DB 299 CGGTGAGCTCTACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 358
QY 1746 GCGCCGAGCAGACCGCCGCTTCCCATAGCTGTCTCATGAGGACAGAACTTGACATGT 1805
DB 359 GCGCGGCTTACTTGT 418
QY 1806 CCAGCTGCCCC 1816
DB 419 GCGTGAAGCGCC 429

RESULT 10
US-09-472-880-1
Sequence 1, Application US/09472880
Patent No. 6274333
GENERAL INFORMATION:
APPLICANT: Daniel CAPUT
APPLICANT: Pascale CHALON
APPLICANT: Pascual FERRARA
APPLICANT: VITA NATALIO
TITLE OF INVENTION: Type 2 Neurotensin Receptor
TITLE OF INVENTION: (hnt-R2)
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh Street
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,880
FILING DATE: 28-Dec-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 9723204
FILING DATE: 17-MAR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,049

INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 1575 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 37..1266
SEQUENCE DESCRIPTION: SEQ ID NO: 1
IS-09-472-880-1

```

	Query Match	Similarity	1.7%	Score 45.4	DB 4	Length 1575
	Best Local	Similarity 46.6%	Pred. No. 0.29			
	Matches 145	Conservative 0	Mismatches 166	Indels 0	Gaps 0	
QY	1506	GGCAGCGCTCCAAAGATCTGGGTCAATGAGGAGCCAAAGCTGGTGTACTTCCAGGGCCACCA	1565			
Db	119	GCCCTGTGGGGCAAGGTGGTGTTCACCGGGCTCTACGCACTCATCTGGGGGTGGGGCGGG	178			
QY	1566	AGGACACGCCCGCTGGAGACACCACTCTACGTGTGTAAGTATGAGGCGCGGCGAGATCG	1625			
Db	179	CGGGCAATGACGCTGTCCGTCGACGTGTGTCGTAAGGCGCGGGCGCGGGCGCGCGCGCG	238			
QY	1626	TACGCGTTCACCAAGCGCGCGGCTCTCCCAATAGCTGTGTCATATGAGCCAGACTTGTGCATATCT	1685			
Db	239	TGCGCCACACAGCTGTCTACGCTCGGCGCTCGGGGCTGCTGTGCTGTGTGGGGGTGC	298			
QY	1686	TGCTGAGCCACTACACGACGCTGAGACAGCCGCCCTCTGCTGCACGCTCTCAACAGCTGAGCG	1745			
Db	299	CGGTGAGACTCTACAGACTTCGTGTGTGGTTCACCTATACCCCTGTGGGCTCTTGGCGACCTGGGCT	358			
QY	1746	GCCCCGACACGACGCCCTTCGCACAAGACGACCCCGCTTCTGGGCTGTAGCATGATGAGAGCGAG	1805			
Db	359	GCCGCGGCTACTACTTCTGTCACAGAGCTGTGGCTACGCCACAGGAGCTGTGAGCGGTGCGAG	418			
QY	1806	CCAGCTGGCCCC	1816			
Db	419	GCCGTGAGCGCC	429			

RESULT 11
 US-07-945-283-1/c
 Sequence 1, Application US/07945283
 Patent No. 5352596
 GENERAL INFORMATION:
 APPLICANT: Cheung, Andrew K.
 APPLICANT: Wesley, Ronald D.
 TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
 TITLE OF INVENTION: Involving The EP0 and LTR Genes
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Curtlis P. Ribando
 STREET: 1815 No. 5352596th University Street
 CITY: Peoria
 STATE: IL
 COUNTRY: USA

```

ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext.513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

```

Query Match	1.7%	Score 45.4	DB 1	Length 8438
Best Local Similarity	48.6%	Pred. No. 0.44		
Matches	124	Conservative	0	Mismatches 131
				Indels 0
				Gaps 0
QY	14	GCCACACCCGGGACCCCAACAGGGCCGACCGAGCGACGACGCCGCCACAGATACCCGGCC	73	
Db	5511	GCCTGGGGCCCCGGGCCCCCTCGGCCCTCTCCGCTCCGCTCCGACCTCTCTCCGCCCGGGGGC	5452	
QY	74	GCCGCTTCCAGGTGCACAGACACTCGTGTGGACGGGCTTCGGAGATCATTCACGGGACGC	133	
Db	5451	TCCTCGGGCCCCGGGGCCCGGAGATCGCCGGCCCCCGGGAGGAAAGCGCGCTCCACCAAC	5392	
QY	134	CGCAGTACTCGGGGCTTATTGTCAACAAGGCGCCCCGACGACTTCAGTTTGTGCAGAG	193	
Db	5391	AACCACTCTCTGCTATGTGGCCGACGCGGGCCCCCGCGACGAGCGGGCCGCTGCTCAACCCG	5332	
QY	194	ACGGATGAGTGTGGGGCCCTCCACATCCACGGCTCTCTACTGTGGAGATGCATATAGCAGC	253	
Db	5331	CTCGGGGAGGCCCTGGCCCGGGCTCCGACCTCTCGGGCGAAGGGCGGCTCCGTACAGGGGC	5272	
QY	254	CGAGAGAACTCCCTC	268	
Db	5271	GCCTGGGACTCTCCGC	5257	

RESULT 12

QY 1749 CCGACGACGACCCCTGCACAA 1770
| | | | | | | | | |
Db 1643 GCGGCGGACCCCGTGACAA 1664

Search completed: December 12, 2002, 10:38:49
job time : 121 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 12, 2002, 10:38:52 ; Search time 21.5 Seconds

(Without alignments)
10097.074 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 4795

Sequence: 1 caagcttaccatgcccaccca.....tctgagcgccgagatccg 2617

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/USPRO.spool/US09976674/runat_04122002_162359_5975/app-query.fasta.1.2759
-DB=Swissprot_40 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09976674 -CCGN_1_1_1_1 -t=unat_04122002_162359_5975 -NCPD=6 -ICPD=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505.5	10.5	765	1	DP44_BOVIN
2	498.5	10.4	767	1	DP44_BOVIN
3	489.5	10.2	766	1	DP44_HUMAN
4	486	10.1	765	1	DP44_FELCA
5	481	10.0	760	1	DP44_MOUSE
6	452	9.4	865	1	DP66_HUMAN
7	447.5	9.3	859	1	DP66_RAT
8	442.5	9.2	863	1	DP66_BOVIN
9	438.5	9.1	761	1	SEPR_MOUSE
10	432.5	9.0	818	1	DAP2_YEAST
11	426	8.9	760	1	SEPR_HUMAN
12	417	8.7	931	1	DAP1_YEAST
13	359.5	7.5	1367	1	AMYH_YEAST
14	279.5	5.8	5179	1	MUC2_HUMAN
15	265.5	5.5	660	1	YHL1_EBV
16	263	5.5	660	1	YHL1_EBV
17	251.5	5.2	1763	1	CA24_ASCSU
18	250.5	5.2	1324	1	IRS2_HUMAN

19	249.5	5.2	1690	1	CA44_HUMAN	P53420	homo sapien
20	249	5.2	657	1	YUXL_BACSU	P39839	hacillus su
21	247.5	5.2	1603	1	CA0F_HUMAN	O07092	homo sapien
22	244.5	5.1	1251	1	Y003_CAEEL	O09350	caenorhabdi
23	242.5	5.1	1262	1	CA13_CHICK	P12105	gallus gall
24	241.5	5.0	1199	1	N121_RAT	P52591	rattus norv
25	239.5	5.0	1783	1	RAA3_CHLRE	O9fec4	chlamydomon
26	238.5	4.9	1464	1	CA13_MOUSE	P08181	mus musculu
27	238.5	4.9	1466	1	CA13_HUMAN	P02461	homo sapien
28	238	4.9	1262	1	CA13_CHICK	P12105	gallus gall
29	238	4.9	1459	1	CA12_MOUSE	P28481	mus musculu
30	237.5	4.9	1049	1	CA13_BOVIN	P04258	bos taurus
31	237.5	5.0	1183	1	DRPL_RAT	P54258	rattus norv
32	235.5	4.9	1366	1	CA21_HUMAN	P08183	homo sapien
33	235	4.9	1461	1	IE18_PRIVF	P11675	pseudorabie
34	235	4.9	1464	1	CA11_HUMAN	P02452	homo sapien
35	232.5	4.8	1460	1	CA11_CANFA	O9xsj7	canis famli
36	232	4.8	1185	1	DRPL_HUMAN	P54259	homo sapien
37	230	4.8	2944	1	CA17_HUMAN	O02368	homo sapien
38	229.5	4.8	1372	1	CA21_RAT	P02466	rattus norv
39	229.5	4.8	1453	1	CA11_CHICK	P02457	gallus gall
40	229	4.7	1418	1	CA12_HUMAN	P02458	homo sapien
41	229	4.7	1453	1	CA11_MOUSE	P11087	mus musculu
42	228	4.8	671	1	CA11_RAT	P02454	rattus norv
43	228	4.8	1446	1	IE18_PRIVKA	P33479	pseudorabie
44	227.5	4.7	1255	1	MUCL_HUMAN	P15941	h mucin 1 p
45	227.5	4.7	1453	1	CA11_CHICK	P02457	gallus gall

ALIGNMENTS

RESULT 1

DP44_BOVIN
AC DP44_BOVIN STANDARD: PRT: 765 AA.

ID P81425; Q8WMC8;
DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Diethylidyl peptidase, IV (EC 3.4.14.5) (DPP IV) (T-cell activation

DE antigen CD26) (adenosine deaminase complexing protein) (ADCP-I)

DE (Activation molecule 3) (ACT3) (WC10).

GN DP44 OR CD26.

OS Bos taurus (Bovine);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=Lymphocytes;

RX PubMed-12073152;

RA Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J.,

RA Bohach G.A.;

RT "Molecular characterization of bovine CD26 upregulated by a

RT staphylococcal superantigen";

RT Immunogenetics 54:216-220(2002).

RL [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.

RC TISSUE=Thymus;

RX PubMed-11981836;

RA Gliddon D.R., Howard C.J.;

RT "CD26 is expressed on a restricted subpopulation of dendritic cells in

RT vivo";

RT Eur. J. Immunol. 32:1472-1481(2002).

RL [3]
RP SEQUENCE OF 1-24.

RC TISSUE=T-cell;

RX PubMed-11598101;

RA Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K.,

RA Naessens J., Bohach G.A.;

RT "Identity of activation molecule 3 on superantigen-stimulated bovine

RT cells is CD26";

RT Infect. Immun. 69:7190-7193(2001).

FN	[4]	SEQUENCE OF 537-546.	
RP		TISSUE=Kidney;	
RC		MEDLINE=98293306; PubMed=9629661;	
RX		Ben-Shooshan I., Parola A.H.;	
RA		"The Cp-I subunit of adenosine deaminase complexing protein from calf kidney is identical to human, mouse, and rat dipeptidyl peptidase IV.";	
RT		Comp. Biochem. Physiol. 119B:289-292(1998).	
RL		-I- FUNCTION: Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. Binds and regulates the activity of ADa.	
CC		-I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb- -Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.	
CC		-I- SUBUNIT: Homodimer.	
CC		-I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in a soluble form (By similarity).	
CC		-I- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and several immune system tissues.	
CC		-I- PPM: The soluble form (SDPP) derives from the membrane form (MDPP) by proteolytic processing (By similarity).	
CC		-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9P.	
CC		-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).	
CC		-----	
CC	EMBL; AF461806; AAL67836.1; -		
DR	EMBL; AY056834; AAL23628.1; -		
DR	MEROPS; S09.046; -		
DR	InterPro: IPR002469; DPPV.N term.		
DR	InterPro: IPR001375; Peptidase_S9.		
DR	InterPro: IPR002471; Prol_endopep_ser.		
DR	InterPro: IPR000379; Ser_estrs_site.		
DR	Pfam; PF00930; DPPV.N term; 1.		
DR	Pfam; PF00326; Peptidase_S9; 1.		
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.		
KW	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;		
KW	Transmembrane; Glycoprotein; Signal-anchor		
KW	CHAIN	1 765	
FT	CHAIN	38 765	
FT	FT		
FT	DOMAIN	1 6	
FT	TRANSMEM	7 29	
FT	FT		
FT	DOMAIN	30 765	
FT	ACT_SITE	629 629	
FT	ACT_SITE	707 707	
FT	ACT_SITE	739 739	
FT	CARBOHYD	84 84	
FT	CARBOHYD	91 91	
FT	CARBOHYD	149 149	
FT	CARBOHYD	218 218	
FT	CARBOHYD	228 228	
FT	CARBOHYD	271 271	
FT	CARBOHYD	280 280	
FT	CARBOHYD	320 320	
FT	CARBOHYD	392 392	
FT	CARBOHYD	495 495	
FT	CARBOHYD	684 684	
CC	SEQUENCE	765 AA; 88369 MW; E2215421F43E116 CRC64;	

Alignment Scores:	2.41e-
Pred. No.:	505.50
Score:	39.15%
Percent Similarity:	

Length:	765
Matches:	180
Conservative:	98

Query Match:	Best Local Similarity:	25.35%	Mismatches:	269
DB:	1	10.548	Indels:	163
			Gaps:	28
US-09-976-674-4 (1-2617) x DPP4_BOVIN (1-765)				
QY 614	TCCTTCATCATATACAGCCAGCTGTGGTGGCCCAATCATCGAGACAGCGAGCGCGG	673		
Db 164	AlaIyValIrrpmsnAsnAspIleIyValIyAsnIgluProAsnSerProserIalnrg	183		
QY 674	CTGACCTTTCGCCACCAAGGTTTATCCATGTCCTGGATGCCACCCAGTGTGGGTG	733		
Db 184	IleThrTrp-----ThrlYlYlSAspValIleTyAsn-----GlyIle	197		
QY 734	GCCACCTTCCTGCATACAGGAAG-----TTCGACCGCTTACAGGGTACTGCTGGTCCC	790		
Db 198	ThrasprIvalIyrgIugIgluIvalPheSerIalTySerIalLeuIrrpIrrSerPro	217		
QY 791	ACACCCCTCCGGGGAAGGTTGAGAGGGCCTCAGACGCTCGAATACCTGATAGCAATC	850		
Db 218	AsnSerThrPhe-----LeuAlaTyrlaIglInPhe	227		
QY 851	GATGAGTTCAGAGGTGGAGTCACTAC-----GTCCCTTCCTGCCTAGACAGAA	901		
Db 228	AsnAspIrrIvalProIeuIleIgluTySerPheTySerAspIuSerIeuIlnTy	247		
QY 902	AGGAAGAGGACATCGTATCGTACCCAGACAGACAGAAATCCCAAGTTGCTTG	961		
Db 248	ProIySthrValIySleIrrProIyrrProIySAlaIyAlaIvalAsnProthIle-----	265		
QY 962	AAATGGCTAGTTCACAGACTGACAGCCAGGCGAAGATCTCTGACCCAGCAAGAGAG	1021		
Db 266	LysPhePheValIalAsnIleSerSerIeuSerProAsnIleAsnAlaThrSerIglIn	285		
QY 1022	CTGTGTGAGGCTTCACCTGCTGTGTCCGAAGGTGAGTATGATCCAGGGCGGGTG	1081		
Db 286	IleIalProIrrIglSerValIleuIleIyAsp--HisTyIleuCyAspValIThrTrp	304		
QY 1082	ACCCGGAGGCAATACGCTTGGGCGCATTTCTTGAGACGGGCCACAGCATGGCTCAG	1141		
Db 305	ValIrrIgluIalIrrIleSer-----LeuIglInIrrPheu---	315		
QY 1142	CTGTCTCTCTCCCGCGGCCCTGTCATCCGAGACAGAAATGAGAGAGCGGCTA	1201		
Db 315	-----	315		
QY 1202	GCCCTGTCAGAGCTGTCCCCAGGAATGTCCAGCCGTATGTGTG-----	1246		
Db 316	-----ArgIrgIleIglInAsnTySerIleMetAspIleCyAsp	328		
QY 1247	TACGAGAGGTCAACCAACGTGTGGATCAATGTTTCATGACATCTTTATTCCTTCCCCAA	1306		
Db 329	TyrAspIrrSerThrIglIyrrIleSer-----	338		
QY 1307	TCAAGAGGAGAGACGAGCTGTCTTTCGCGGCCCAATGATGCAGACCGGCTTTCG	1366		
Db 339	-----SerValIglIyrrIglIn	343		
QY 1367	CATTGTACAAATGACACCGCGCTTTTAAATCCAGAGGCTACGATGTGAGTGAAGCCCTTC	1426		
Db 344	HisIleIuIleSerThrTrp-----GlyTrpValIglIyrrPhe	356		
QY 1427	AGCCCGGGGGAATGAATTT-----AAGTGGCCCATTTAAG	1482		
Db 357	ArgProIalIuProIrrHisPheThrSerAspIglIyAsnSerPheTyTrIyIleIleSerAsn	376		
QY 1463	GAAGAG-----ATTGCTCTG	1477		
Db 377	GlugIugIyTrIyTrIyShIleCyShIlePheIglInThrAspIyArgAsnCySerThrPheIle	396		
QY 1478	ACCAAGCGTGAATGGAGGTTTGGCGGAGCACGGCTCCAGATCTGGGTCAATGAGAG	1537		
Db 397	ThrIySglIyAlaIrrpIglIyAlIleIyIleIgluAlaIeuThrSer	411		


```

DR MEROPS; S09.003; -.
DR Interpro: IPR002469; DPIP_N_term.
DR Interpro: IPR001375; Peptidase_S9.
DR Interpro: IPR002471; Prol_endopep_ser.
DR Interpro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Pfam: PF00930; DPIP_N_term; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
DR HydroLase: AminoPeptidase; Dipeptidase; Serine protease;
KM Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 767 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT CHAIN 37 767 (MDP).
FT CHAIN 281 767 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT CHAIN 281 767 (SDP).
FT CHAIN 281 767 DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE
FT CHAIN 281 767 FORM.
FT TRANSMEM 7 28 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSMEM 7 28 (POTENTIAL).
FT DOMAIN 29 767 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 631 631 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 709 709 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 741 741 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 A -> R (IN REF. 2).
FT CARBOHYD 183 183 I -> T (IN REF. 2).
FT CARBOHYD 332 332 T -> N (IN REF. 2).
FT CARBOHYD 352 352 C -> V (IN REF. 2).
FT CARBOHYD 394 394 V -> D (IN REF. 2).
FT CARBOHYD 362 362 L -> F (IN REF. 2).
FT CARBOHYD 624 624 R -> Q (IN REF. 2).
FT CONFLICT 767 AA: 88003 MW: AA17C0BC6F0E4652 CRC64;
SQ SEQUENCE

```

```

Alignment Scores:
Pred. No.: 5,41e-18 Length: 767
Score: 498.50 Matches: 210
Percent Similarity: 37.96% Conservative: 121
Best Local Similarity: 24.08% Mismatches: 328
Query Match: 10.40% Indels: 213
DB: 1 Gaps: 38

```

US-09-976-674-4 (1-2617) x DRP4_RAT (1-767)

```

OY 197 GATGAGTCTGGCCGCCACCTCCACCGCTCTAC-----TACCTGGGAATGCCA 244
    |||||::: ||| ||| ||| |||||
Db 31 ASGGLAALALALALASPSERIALARGTHTYTHLEUALASPRYLEULYASANTHR 50
OY 245 TATGGACGCCGAGAGACTCCCTCTTAC--TGTGAGATTCCCAAGAGTCCGGAAA 301
    ::::: ||||| ::: |||||
Db 51 PHEATVALLYSERYSERILEUARGTRVALSERASPERSERGLUTRYLEUTRYLSGLN 70
OY 302 GAGGCTCTGCTGCTCTGCTGCTGGAAGACATCTGATCATTTCCAGGCCACCCCCAC 361
    ||| ::::| ||||| -----PheasnaLa-----GLu 80
Db 71 GLUASNAASNLLEULEU-----PheasnaLa-----GLu 80
OY 362 CATGGAGTCTACTCTCGGAGGAGAGCTGAGGAGCGGAACGCCCTGGGGCTTTC 421
    ||||| ::: ||||| |||||
Db 81 HISGLY-----AsnSerSerIlePheLeuGLuAsnSerThrPheGLuIlePhe 96
OY 422 GGCATCACTCTACACTCTCAC---AGCGAGATGCGCTTCTCTTCCAGGCCAGC 478
    ||| ::: ||||| |||||
Db 97 GLYASPSERILESERASPRYSERVALSERPROASPRARGHEUHPHEVALLEULEUGLUTYR 116
OY 479 AACAGCCTCTTCCACTGTGTGCGACGGCGGACAGACAGCGCTTCACTGGTGTCCCTATGAAA 538
    ||| ::: |||
Db 117 ASNTYVALLYLSGLINTPRAPHISSETYTRYTHRALASERTYISERILETRYASPLEUASN 136

```

```

OY 539 CCGSTGGAATCAAGACCCAGTGTCTGAGGGCCCGGATGAGACCCCAAAATCTCCCTGCC 598
    ::::: ||||| |||
Db 137 LYSARGGLNLEULERTHRLGULYSLIETPROASNASTHCLINTRPLETRTPSER 156
OY 599 GACCTGCGCTTC--TTTCCTTCATCAATACAGCGACCTGTGGTGGCCACATCGAG 655
    ::::: |||||
Db 157 GLNGLGLYHISLYSLEULATYTRVALTRPYSASNAASPILETYRVALLYSILEGLUPRO 176
OY 656 ACAGGGAGAGAGGGGGAGCTTCTGCGACCAAGATTTATCAATGATGCTGATGAC 715
Db 177 HISLEUPROSERHISARGILETHR-----SerThGLYLSGLIUSNVALLIHPHEN 194
OY 716 CCCAAGTCTGGGGTGTGGCCACCTTGTGATCATACAGAGAG---TTGACAGCGCTTCACT 772
    ||||| ::::| |||||
Db 195 -----GLYLSASNAAPTRVALTRYGLIUSGLNLIETHPHEGLYALATYR 210
OY 773 GGGTACTGTGTGTCGCCACAGCTCTGGGAAGTTCAAGAGGGCTCAAGACGCTGCA 832
    ||||| |||
Db 211 ALALEUTRTPSERPROASNLTYTHRPE----- 220
OY 833 ATCSTGATGAGAGATGCATGATGATGCGAGGTGAGGTCATTTCAC-----GTCCCC 883
    ::: ||| ::: |||||
Db 221 LEUALATYRALAIPHEASNAAPTRHGLYVALPROLEULIETUTYRSETPHETYSER 240
OY 884 TCTCTGCGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
    ::::: |||||
Db 241 ASPGLISERLEULINTRYPROLYSTHTRVALTRIPLETRYRPROLYSALGLYALVAL 260
OY 944 AATCCCAAGATTGCTTGAATAGCTGCTGAGTTCACACTGACAGCCAGGCAAGATCTGC 1003
    ::::: |||||
Db 261 ASNPTRHTRVAL-----LYSPHEPHEILEVALASPTHASPERSERLEUSETRTHR 278
OY 1004 TGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063
    ::::: |||||
Db 279 TRHTRHILEPROMETCLINLETHTALAPRO---ALISERVATHRTRHGLYASPHISTYR 297
OY 1064 ATCGCAGAGGGCGGTGAGACCCGAGTGCMAATAGCCCTGGGCCATGTCTGAGACCG 1123
    ::: ||| ::: |||
Db 298 LEUCYASPRVALALATRPALSERGLUASPARGLISER----- 310
OY 1124 CCCACAGTGTGCTCAGCTCGCTCTCTCCCGCCGCTTTCATCCGAGACAGAG 1183
    ||||| |||
Db 311 ---LeuGLNTRPLeu----- 314
OY 1184 AATGAGAGAGAGGGGTAGGCTGTGCTGCGACAGAGTGTCCGAGAAATGCCAGTATG 1243
    ||| ::::| |||
Db 315 -----ArgArgIleGLIAsnTYR 321
OY 1244 GTG-----TACGAGAGAGTCCACAGCTGTGATCAATGTTCATGACATC 1288
    ||| ||| ::::: ||| |||||
Db 322 VALMETALALECYSASPRYTRASPRLYSTHTRHLEUVALTRP----- 335
OY 1289 TTCTATTCCTCCCGCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1348
    ::::: |||||
Db 336 -----Asn 336
OY 1349 TGCAGACGGGCTTTCGCAATTGTACAAAGTACAGCGCGTTTAAATGCCAGGCTTAC 1408
    ||| ||| |||
Db 337 CYSPTHTHTRHGLNLIUS-----IleGLUTHRSERALATHR 349
OY 1409 GATTGAGTGAAGCCTTTCAGCCCGGGGAGAGATGAATTTAAG----- 1450
    ||| ||| |||
Db 350 GLYTRPCYGLYARGPHEATRPROLAIGLUPRONHISPHETHSERASRLYSERSEPH 369
OY 1451 -----TGCSCCATTAAGAGAG--- 1468
    ||| |||
Db 370 TYRGLYILEVALSERASPRYASPRGLYTRYLYSHISILECYSGLINPHEGLIUSASPAR 389
OY 1469 -----ATGCTCTGACAGCGGAGATGGAGAGGTTTGGGAGAGCAG 1510
    ::::: |||||
Db 390 LYSPTROGLINVALCYSTHTRPHEILETHRLYSGLYALATRPGLIUSALLESER----- 407

```



```

QY 1511 GGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTTACTTCCAGGGCACG---AAG 1567
Db      ||| ||| |||
QY 408 -----IleGuaIaleuThiSerAspTyrTyrTyrIleSerAsnGluTyrLys 424
QY 1568 GACACGCCCTGGAGACCAACCTCTACGTGGTACGTATGAGCGCGCGAGATCGTA 1627
Db      ||| ||| ||| ||| ||| ||| |||
QY 425 GluMetProGluGlyArgAsnLeuTyrLysIleGln----- 436
QY 1628 CGGCTCACACGCCCGGCTTCTCCATAGCTCCATGAGCCGAACTTCGACATGTC 1687
Db      ||| ||| ||| ||| ||| ||| |||
QY 437 -----LeuThiAspHisThrAsnLysLysCys---LeuSerCysAspLeuAsnProGlu 453
QY 1688 GTCAGCCACTACAGCAGCAGCTGAGCAGCCCGCTGCGACAGCTCTCAACAGCTAGC--- 1744
Db      ||| ||| ||| ||| ||| ||| |||
QY 454 ArgCysLysIleTyrTyrSerValSerLeuSerLysLysLysLysLysLysLys 473
QY 1745 ---GGCCCC----- 1750
Db      ||| ||| ||| ||| ||| ||| |||
QY 474 ArgGlyProGlyLeuProLeuTyrThrLeuHisArgSerThrAspGlnLysGluLeuArg 493
QY 1751 -----GAGCAGCAGCCCTGCACAGCAGCCCGCTTCTGGGCTAGCATGATGAG 1801
Db      ||| ||| ||| ||| ||| ||| |||
QY 494 ValLeuGluAspAsnSerAlaLeuAspLys-----MetLeuGln 506
QY 1802 GCAGCCAGCTGCCCCCGGATTATGTTCTCCAGAGATCTTCATTTCCACAGCCGCTCG 1861
Db      ||| ||| ||| ||| ||| ||| |||
QY 507 AspValGln-----MetProSerLysLysLeuAspPheIleValLeuAsn 521
QY 1862 GATGTGGGCTC---TAGCGCATGATCTACAGCCCGCTTGCAGCCAGGAGAAAG 1918
Db      ||| ||| ||| ||| ||| ||| |||
QY 522 GluThrArgPheTyrGlyGlnMetIleLeuProHis---PheAspLysSerLysLys 540
QY 1919 CACCCACCGCTCTTGTATATGAGGCCCGCCAGGTGACGTGGTGAATAACTCTTC 1978
Db      ||| ||| ||| ||| ||| ||| |||
QY 541 TyrProLeuLeuIleAspValTyrAlaGlyProCysSerGlnLysAlaAspAlaPhe 560
QY 1979 AAAGGCATCAAGTACTGCGGCTCAAC-----ACATGCGGCTCCCTGGCTAGCGCC 2029
Db      ||| ||| ||| ||| ||| ||| |||
QY 561 -----ArgLeuAsnThrAlaThrTyrLeuAlaSerThrGluAsnIle 574
QY 2030 GTGGTTGTG---ATTGACGGCAGGGGCTCTGTACGCGAGGCTTCGGTTGGAAGGGCC 2086
Db      ||| ||| ||| ||| ||| ||| |||
QY 575 IleValAlaSerPheAspGlyArgGlySerGlyTyrGlnGlyAspLysIleMetHisAla 594
QY 2087 CTGAAAACCAAAATGGCGCAGGTGAGATGAGAGACCAAGCGGTTCACAGTGGCCATC 2143
Db      ||| ||| ||| ||| ||| ||| |||
QY 595 IleAsnLysArgLeuGlyThrLeuGluValGluAspGlnIleGluAlaAlaArgGlnPhe 614
QY 2144 GTGCGCCGAGAGTATGGCTTCATGACCTGAGCCGAGTGGCATTCAGCTGCTCTAC 2203
Db      ||| ||| ||| ||| ||| ||| |||
QY 615 Leu-----LysMetGlyPheValAspSerLysArgValAlaIleTyrPglTyrSerTyr 632
QY 2204 GGGGGCTTCTCTCGCTCATGCGGCTAATCCACAAGCCCGAGTTCACAGTGGCCATC 2263
Db      ||| ||| ||| ||| ||| ||| |||
QY 633 GlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValPheLysCysGlyIle 652
QY 2264 GCGGGGCGCCCGGTACCGCTGTGATGAGCCATGACACAGGTTACAGTACCGCTACTG 2323
Db      ||| ||| ||| ||| ||| ||| |||
QY 653 AlaValAlaProValSerArgTyrPglTyrTyrAspSerValTyrThrGluArgTyrMet 672
QY 2324 GACGTC-----CCTGAGAACACACAGCAGCGCTATGAGGGGGTTCGGCTCGCCTGCAC 2377
Db      ||| ||| ||| ||| ||| ||| |||
QY 673 GlyLeuProThrProGluAspAsnLeuAspHisThrArgAsnSerThrValMetSerArg 692
QY 2378 GTGAGAAAGCTGCCCAATGAGCCCAACCGCTTGTATCTCCACAGGCTTCTTGACGAA 2437
Db      ||| ||| ||| ||| ||| ||| |||
QY 693 AlaGluAsnPhe-----LysGlnValGluTyrLeuLeuIleHisGlyThrAlaAspSpr 710
QY 2438 AACGTGACCTTTTCCACACAAACTTCCTGTCGCCAAGATCCGAGCGGAGAAACT 2497
Db      ||| ||| ||| ||| ||| ||| |||
QY 711 AsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAspAlaGlyValAsp 730
QY 2498 TACCAGCTCCAGATCTACCCCAAGAGACACAGATATTCGCTCCCGAGTGGGCGAG 2557

```

```

Db 731 PheGlnAlaMetTyrTyrThrAspGluAspHisGlyIleAlaSerSerThrAlaHisGln 750
QY 2558 CASATGAAAGTCAAGCTTGTCTGCACCTTCTTACAGGAA 2593
Db 751 HisIleTyrSerHisMetSerHisPheLeuGlnGln 762
RESULT 3
DPP4_HUMAN
ID DPP4_HUMAN STANDARD; PRT; 766 AA.
AC P27487;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (TP103) (Adenosine deaminase complexing protein-2)
DE (ADABP).
GN DPP4 OR ADCP2 OR CD26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92329551; PubMed=1352704;
RA Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
RT "Molecular cloning and sequence analysis of human dipeptidyl
RT peptidase IV, a serine proteinase on the cell surface.";
RL Biochim. Biophys. Acta 1131:333-336(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95012434; PubMed=7927537;
RA Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
RT "Genomic organization, exact localization, and tissue expression of
RT the human CD26 (dipeptidyl peptidase IV) gene.";
RL Immunogenetics 40:331-338(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Periphereal blood;
RX MEDLINE=93171637; PubMed=8094732;
RA Tanaka T., Camerlin D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
RA Dahlberg H.N., Schlossman S.F., Morimoto C.;
RT "Cloning and functional expression of the T cell activation antigen
RT CD26.";
RL J. Immunol. 149:481-486(1992).
RN [4]
RP ERRATUM.
RX MEDLINE=93171637; PubMed=8094732;
RA Tanaka T.;
RL J. Immunol. 150:2090-2090(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 1-551 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=92165847; PubMed=1347043;
RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
RA Trottet P., Barbat A.;
RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
RT colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
RT human coding sequence and changes of dipeptidyl peptidase IV mRNA
RT levels during cell differentiation.";
RL J. Biol. Chem. 267:4824-4833(1992).
RN [7]
RP SEQUENCE OF 545-766 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=91024044; PubMed=1977364;
RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;

```

```

RT      "Isolation of a cDNA probe for the human intestinal
RT      dipeptidylpeptidase IV and assignment of the gene locus DPP4 to
RT      chromosome 2."
RL      Ann. Hum. Genet. 54:191-197(1990).
RM      [8]
RP      SEQUENCE OF 1-31 FROM N.A.
RX      MEDLINE=96067599; PubMed=7487939;
RA      Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
RT      "Human dipeptidyl peptidase IV gene promoter: tissue-specific
RT      regulation from a TATA-less GC-rich sequence characteristic of a
RT      housekeeping gene promoter."
RL      Biochem. J. 311:835-843(1995).
RM      [9]
RA      PARTIAL SEQUENCE.
RP      TISSUE=Kidney;
RC      MEDLINE=93210468; PubMed=8096237;
RX      Morrison M.E., Vijayasaradhi S., Engelstein D., Albino A.P.,
RA      Houghton A.N.;
RT      "A marker for neoplastic progression of human melanocytes is a cell
RT      surface ectopeptidase."
RL      J. Exp. Med. 177:1135-1143(1993).
CC      -I- FUNCTION: Removes N-terminal dipeptides sequentially from
CC      polypeptides having unsubstituted N-termini provided that the
CC      penultimate residue is proline. Plays a role in T cell activation.
CC      -I- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-1-
CC      xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC      Xcc is neither Pro nor hydroxyproline.
CC      -I- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
CC      -I- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC      a soluble form.
CC      -I- PTM: The soluble form (SDP) derives from the membrane form (MDPP)
CC      by proteolytic processing.
CC      -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC      -I- DATABASE: NAME=PROM; NOTE=CD guide CD26 entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U13735; AAB60646.1; JOINED.
DR      EMBL; U13710; AAB60646.1; JOINED.
DR      EMBL; U13711; AAB60646.1; JOINED.
DR      EMBL; U13712; AAB60646.1; JOINED.
DR      EMBL; U13713; AAB60646.1; JOINED.
DR      EMBL; U13714; AAB60646.1; JOINED.
DR      EMBL; U13715; AAB60646.1; JOINED.
DR      EMBL; U13716; AAB60646.1; JOINED.
DR      EMBL; U13717; AAB60646.1; JOINED.
DR      EMBL; U13718; AAB60646.1; JOINED.
DR      EMBL; U13719; AAB60646.1; JOINED.
DR      EMBL; U13720; AAB60646.1; JOINED.
DR      EMBL; U13721; AAB60646.1; JOINED.
DR      EMBL; U13722; AAB60646.1; JOINED.
DR      EMBL; U13723; AAB60646.1; JOINED.
DR      EMBL; U13724; AAB60646.1; JOINED.
DR      EMBL; U13725; AAB60646.1; JOINED.
DR      EMBL; U13726; AAB60646.1; JOINED.
DR      EMBL; U13727; AAB60646.1; JOINED.
DR      EMBL; U13728; AAB60646.1; JOINED.
DR      EMBL; U13729; AAB60646.1; JOINED.
DR      EMBL; U13730; AAB60646.1; JOINED.
DR      EMBL; U13731; AAB60646.1; JOINED.
DR      EMBL; U13732; AAB60646.1; JOINED.
DR      EMBL; U13733; AAB60646.1; JOINED.
DR      EMBL; U13734; AAB60646.1; JOINED.
DR      EMBL; M74777; AAS51943.1; -.
DR      EMBL; BC013329; AAH13329.1; -.
DR      EMBL; M80536; AAA52308.1; -.

```

Accession	Protein Name	Length (aa)	Score	Similarity (%)	Best Local Similarity (%)	Query Match (%)	DB
DR EMBL: X60708; CAA43118.1; -	EMBL: X60708; CAA43118.1; -	766	1.52e-17	489.50	36.27%	22.43%	DB
DR EMBL: 579876; AAB35614.1; -	EMBL: 579876; AAB35614.1; -	766	1.52e-17	489.50	36.27%	22.43%	DB
DR PIR: S24313; CDH026.	PIR: S24313; CDH026.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR MEROPS: S09.003; -	MEROPS: S09.003; -	766	1.52e-17	489.50	36.27%	22.43%	DB
DR Genew: HGNC:3009; DPP4.	Genew: HGNC:3009; DPP4.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR MIM: 102720; -	MIM: 102720; -	766	1.52e-17	489.50	36.27%	22.43%	DB
DR InterPro: IPR002469; DPPV_N.term.	InterPro: IPR002469; DPPV_N.term.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR InterPro: IPR001375; Peptidase_S9.	InterPro: IPR001375; Peptidase_S9.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR InterPro: IPR002471; Prol_endopep_ser.	InterPro: IPR002471; Prol_endopep_ser.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR InterPro: IPR000379; Ser_est_sile.	InterPro: IPR000379; Ser_est_sile.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR Pfam: PF00326; Peptidase_S9; 1.	Pfam: PF00326; Peptidase_S9; 1.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR Pfam: PF00930; DPPV_N.term; 1.	Pfam: PF00930; DPPV_N.term; 1.	766	1.52e-17	489.50	36.27%	22.43%	DB
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.	PROSITE: PS00708; PRO_ENDOPEP_SER; 1.	766	1.52e-17	489.50	36.27%	22.43%	DB
KM Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;	Hydroxylase; Aminopeptidase; Dipeptidase; Serine protease;	766	1.52e-17	489.50	36.27%	22.43%	DB
KW Transmembrane; Glycoprotein; Signal-anchor.	Transmembrane; Glycoprotein; Signal-anchor.	766	1.52e-17	489.50	36.27%	22.43%	DB
FT CHAIN 1	CHAIN 1	766	1.52e-17	489.50	36.27%	22.43%	DB
FT FT	FT	766	1.52e-17	489.50	36.27%	22.43%	DB
FT CHAIN 39	CHAIN 39	766	1.52e-17	489.50	36.27%	22.43%	DB
FT FT	FT	766	1.52e-17	489.50	36.27%	22.43%	DB
FT DOMAIN 1	DOMAIN 1	6	1.52e-17	489.50	36.27%	22.43%	DB
FT TRANSMEM 7	TRANSMEM 7	28	1.52e-17	489.50	36.27%	22.43%	DB
FT FT	FT	28	1.52e-17	489.50	36.27%	22.43%	DB
FT DOMAIN 29	DOMAIN 29	766	1.52e-17	489.50	36.27%	22.43%	DB
FT ACT_SITE 630	ACT_SITE 630	766	1.52e-17	489.50	36.27%	22.43%	DB
FT ACT_SITE 708	ACT_SITE 708	766	1.52e-17	489.50	36.27%	22.43%	DB
FT ACT_SITE 740	ACT_SITE 740	766	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 85	CARBOHD 85	85	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 92	CARBOHD 92	92	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 150	CARBOHD 150	150	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 219	CARBOHD 219	219	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 229	CARBOHD 229	229	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 281	CARBOHD 281	281	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 321	CARBOHD 321	321	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 520	CARBOHD 520	520	1.52e-17	489.50	36.27%	22.43%	DB
FT CARBOHD 685	CARBOHD 685	685	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 6	CONFLICT 6	6	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 7	CONFLICT 7	7	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 437	CONFLICT 437	437	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 557	CONFLICT 557	557	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 663	CONFLICT 663	663	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 663	CONFLICT 663	663	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 663	CONFLICT 663	663	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 663	CONFLICT 663	663	1.52e-17	489.50	36.27%	22.43%	DB
FT CONFLICT 6							

```

Db      96  AspluInhegIlyHisSerIleasnAspTyrSerIleSerProaspGlyGlnPheIleu 115
OY      470  CAGCCGCAAGAACAGCCCTTCCACTGCGGACGGCGGCAAGACGGTTATGTCG 529
Db      116  LeuGIuTyrSnsYrValLysGIuTyrPArgHisSerYrThrAlaSerYrAspIleYr 135
OY      530  CCTATGAACCGCGTGAATCAAGACCCAGTGCAGGCGCCCGGATGAGCCCAAAATC 589
Db      136  AspleuSnsYrAspGlnIleuIleThrGIuIuArgIleProaspAsnThrGIuTyrVal 155
OY      590  TGCCCTCCGACCCCTGCCCTTC---TTCCTCTTCAATAACAGCAGCCTGTGGGTGCC 646
Db      156  ThrTrpSerProValGIuHisIleuAlaTyrValTrpAsnAsnAspIleTyrValLys 175
OY      647  AACATCAAGACAGCGGAGGAGCGGGCTGACCTTCCGCCCAAGTTATCCAAATGTC 706
Db      176  ILeGIuProaspIleuProSerTyrArgIleThrTrp-----ThrGIuLysGIuAspIle 193
OY      707  CTGATGACCCCAAGTCTGCGGTGTGGCCACCTTGCATACAGGAAG---TTCGAC 763
Db      194  ILeTyrAsn-----GlyIleThrAspTrpValTyrGIuGIuValPheSer 209
OY      764  CGCTTCACTGGTACTGTGTGTGCCACAGCCTCTGGGAAGGTTCAAGGGCCTCAG 823
Db      210  AlatySerAlaIleuTrpTrpSerProaspGlyThrPhe----- 222
OY      824  ACGCTGGCAATCCGTATGAGAAAGTGAATGCCAGATCCAGATGAGATTCAC----- 877
Db      223  -----LeuAlaTyrAlaGlnPheAsnAspThrGIuValProIleGIuTyrSer 239
OY      878  ---GTCCCTCTCTCGCGCTGAGAAAGAGAGAGCGACTGTATCGGTACCCGAGACA 934
Db      240  PheTyrSerAspGIuSerIleuGlnTyrProLysThrValAlaGValProTyrProLysAla 259
OY      935  GGCAGCAAGAATCCCAAGATTCCTTGAACCTGCTGATTCAGACTGACAGCCAGGCG 994
Db      260  GLyAlaValAsnProThrVal-----LysPheValValAsnThrAspSerLeuSer 277
OY      995  AAGATCTCTTCGACCCAGAGAGAGAGAGTGTGACAGCCCTTCACTGCTTCCCGAAG 1054
Db      278  SerAlaThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGIuAsp 297
OY      1055  GTGAGTACATACCGCCAGGCGGTGAGACCGGATGGCAATACGCGCTGGCCATGTC 1114
Db      298  ---HisTyrIleuCyAspValThrTrpAlaThrGlnGIuArgIleSer----- 312
OY      1115  CTGACCGGCGCCAGAGTGGCTCCAGCTGCTCTCCCTCCCGCGCGCTTCATCCCG 1174
Db      313  -----LeuGIuTyrLeu----- 316
OY      1175  AGCACAAGATGAGAGAGACGGCTAGCCCTCTCCAGAGCTGTCCCGAAGTCCAG 1234
Db      317  -----ArgArgIleGln----- 320
OY      1235  CCGATATGTGNG-----TAGAGAGAGTCAACCAAGCTGTGATCAATGTT 1279
Db      321  AsnTyrSerValMetAspIleCysAspTyrAspIleSerGIuArgTyr----- 337
OY      1280  CATGACATCTTCTATCCCTTCCCAATCAAGAGGAGAGAGAGAGCTCTCTTCGCG 1339
Db      337  ----- 337
OY      1340  GCCAATATGCAAGACCGGCTTGTCCATTTGTACAAGTCAACGCCGTTTAAATCC 1399
Db      338  -----AsnCysIleuValAlaArgGlnHisIleGIuMetSerThr----- 351
OY      1400  CAGGGTACGATGTGAGTGAAGCCCTTCAGCCCGGGGAGAGATGA----- 1444
Db      352  -----GlyTyrValGIuArgPheArgProSerGIuProHisPheThrLeuAspGIy 368
OY      1445  -----TTTAAG 1450

```

```

Db      369  AsnSerPheTyrLysIleIleSerAsnGIuGIuTyrArgHisIleCysTyrPheGln 388
OY      1451  TGCCCATTAAGAAAGATGTCCTGACACCGCTGAATGGAGGTTTGCGGAGGAC 1510
Db      389  IleAspLysLysAspCysThrPheIleThrLysGIuTyrTrpGIuValIleGIuIleGIu 408
OY      1511  GGCCTCAAGATCTGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCAACAGGAC 1570
Db      409  AlaIleu-----ThSerAsp 413
OY      1571  ACGCGCTGAGACCAACCTCTACGTGTGACATATGAGGGCGCGG----- 1618
Db      414  -----TyrLeuTyrTrIleSerAsnGIuTyrLysGIuMetProGIuGIy 428
OY      1619  ---GAGATCGTACGCTTACCAACCGCGCTTCCCTACCTGCTCATAGCCAGAC 1675
Db      429  ArgAsnIleuTyrLysIleGlnLeuSerAspTyrThrLysValThrCysLeuSerCysGIu 448
OY      1676  TTCACATGTTGTCAGCCACTACAGCAGCGTGAAGACAGCGCGCGCTGCTGACGTTAC 1735
Db      449  LeuAsnProGIuArgCysGlnTyrTyrSerValSerPheSerLysGIuAlaLysTyrTyr 468
OY      1736  AAGCTG-----ACCGCGCC----- 1750
Db      469  GlnIleuArgCysSerGIuProGIuLysIleuProIleuTyrThrIleHisSerSerValAsnAsp 488
OY      1751  -----GAGACAGACCCCTGCAAAAGACAGCCCGCTTGGGCT 1789
Db      489  LysGIuLeuArgValIleuGIuAspAsnSerAlaLeuAspLys----- 502
OY      1790  AGCATGATGAGGAGGAGCCAGCTGCCCGGATTAATGTTCTCCAGATGTCATTTTC 1849
Db      503  ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
OY      1850  CACACGGCTCGGATGTGCGGCTC---TACGCGATGATCTCAAAAGCCAGCCTTGCAG 1906
Db      517  IleIleuAsnGIuThrLysPheTyrTrpGIuMetIleuProHis---PheAsp 535
OY      1907  CCAGGAAAGAAAGACCCCGCTCTTTGTATATGAGGCGCCAGAGCTGAGTGTG 1966
Db      536  LysSerLysLysTyrProIleuLeuAspValTyrAlaGIuProCysSerGIuLysAla 555
OY      1967  AATACCTCTTCAAAAGCATCAAGTACTGGCGCTCAAC-----ACATGGCCTTC 2017
Db      556  AspThrValPhe-----ArgLeuAsnThrPalaThrTyrLeuAlaSer 569
OY      2018  CTGGGCTACGGCGGTGTGTG---ATTGACGCGAGGGCTCTCTCAGCAGAGGCTTCG 2074
Db      570  ThrGIuAsnIleIleValAlaSerPheAspGIuArgGIuSerGIuTyrGIuLysPlys 589
OY      2075  TTCGAAAGGCGCCCTGAAAACCAATGGGCCAGGTGGAGTGCAGAGCAGGTGAGGCG 2134
Db      590  IleMetHisAlaIleAsnArgArgLeuGIuThrPheGIuValGIuAspGIuIleGIuAla 609
OY      2135  CTG---CAGTTCGTGGCGGAGAGATGAGCTTCAATGAGCAGGCGGTGCAATCAT 2191
Db      610  AlaArgGlnPhe-----SerLysMetGIuPheValAspAsnLysArgIleAlaIleTrp 627
OY      2192  GCGTGTCTTACGGGGGCTCTCTGCTCATGGGCTGAATTCACAAACCCCGAGTGTTC 2251
Db      628  GlyTrpSerTyrGIuGIuTyrValThrSerMetValIleuGIuSerGIuValPhe 647
OY      2252  AAGTGGCCATCGCGGGTGGCCCGGTACCGCTGTGGATGGCTTACAGACAGGTAAC 2311
Db      648  LysCysGIuIleAlaValAlaProValSerArgTrpGIuTyrTyrAspSerValTyrThr 667
OY      2312  GAGGCTCATGAGACGTC-----CTGAGAAACAACGACAGCGGTATGAGGCGGCTCC 2365
Db      668  GIuArgTyrMetGIuLeuProThrProGIuAspAsnLysValArgAsnSerThr 687
OY      2366  GTGGCGCTGACAGTGAAGCTGCCAATGAGGCCAAGCCGCTTATGCTTCACAGCGC 2425
Db      688  ValMetSerArgAlaGIuAsnPhe-----LysGIuValGIuTyrLeuLeuIleHisGIy 705

```



```

QY 799 CTGGAGAGTTCCAGAGGCGCTCAAGACGCTCGAATCCGTATAGAGATCGATGATC 858
   : |||: ||| :|||: |||
Db 311 r-----AlgleuAlaItyrAlaIAlaIleAsnAspSe 321
QY 859 CGAGGTGGAGTGCATTCACGTCCCTCT-----CTCGCGCTAGAGA 900
   : ||| :|||: |||: |||
Db 321 rArGValProIleMetCylLeuProThrItyrThnGlySerIleTyrProThrValys-- 340
QY 901 AAGGAAGACGAGCTCGTATGGTACCCAGAGACAGCAAGATCCCAAGATGGCTT 960
   : ||| :|||: |||: |||
Db 341 -----ProIyHisTyrProIySalGlySerGluAsnProSerIleSerIle 356
QY 961 GAAACTGGTGAAGTCCAGACTGCACAGCCAGCAAGATGCTCCGACCCAGAGAGA 1020
   : ||| :|||: |||: |||
Db 356 uHISValIle-----GlyLeuAsnGlyProThrHisAspLeuG 369
QY 1021 GCTGTGGACCCCTTCAGCTCGCTGTCCGAAAGTGGAG-----TACATCGCGAGGC 1074
   : |||: ||| :|||: |||
Db 369 uMetMetProProAspAsp-----ProArgMetArgGluTyrTyrIleThrMetVa 386
QY 1075 CGGGTGGACCCGGGATGGCAATACGCTGGCCATCTCTGGACCGCGCCAGCACTG 1134
   : ||| :|||: |||: |||
Db 386 lYsTyrAlaIthSerThrIySValAla---ValThrIyLeuAsnArgAlaGlnAsnVa 405
QY 1135 GCTCCAGCTGCTCTCTCCCGCGCCCTGTTCATCCGAGCAGACAGATGAGAGCA 1194
   : ||| :|||: |||: |||
Db 405 lSerIleIleuThrLeuCySAspAlaThrIthrGlyValCySThrIySHisGluAspG 425
QY 1195 GCGGCTAGCCCTGCCAGAGCTGTCCCGAGAAATGTCACCCGTATGGTGTACAGAGA 1254
   : ||| :|||: |||: |||
Db 425 u----- 425
QY 1255 GGTCCACCAAGTGTGATC-----AATGTTATGACATCTTCTATCCCTTCCCCCA 1305
   : |||: ||| :|||: |||
Db 426 ----SerGluAlaIthrLeuHisArgGlnAsnGluIuProValPhe----- 439
QY 1306 ATCAGAGGAGAGAGAGAGTGTGCTTCTCCGCGCCCAATGAATGCAAGACGGCTCTG 1365
   : |||: ||| :|||: |||
Db 440 -SerIySAspGlyArgIySPherPheIleArgAlaIlePro---GlnGlyGlyArgG 458
QY 1366 CCATTTGTCAAAATCACCCCGCTTTTAAATCCAGAGGCTAGCATTTGAGTGAGCCCTT 1425
   : ||| :|||: |||: |||
Db 458 yLysPheIyHisIleThrAl-----SerSerSerGlnProAs 471
QY 1426 CAGCCCGGGGAAAGATGAATTAATGCCCCATTAAAGAAAGATTCCTGACACACGG 1485
   : ||| :|||: |||: |||
Db 471 nSerSerAsnAspAsn-----IleGln-----SerIleThrSerG 483
QY 1486 TGAATGGAGGTTTGGCGAGGACAGCGCTCCAGAGATCTGGGTC---AATGAGGAGACCA 1542
   : |||: ||| :|||: |||
Db 483 yAspTTPAspVal-----ThrIySileuAlaItyrAspGluIySglYAs 498
QY 1543 GCTGGTGTACTTCCAGGACCAAGACAGCGCTGAGACACACCTCAAGTGTGAC 1602
   : |||: ||| :|||: |||
Db 498 nLysIleTyrPheIeuSerThrGluAspLeuProArgArgGlnIleuIySerAlaAs 518
QY 1603 CTATGAGCGCGCGGAGATCTGACGCTCACCAGCGCGGCTCTCCCAT----- 1654
   : |||: ||| :|||: |||
Db 518 nThrGluGly-----AsnPhesAsnArgGlnCysLe 528
QY 1655 -AGCTGTCATGAGCCAGAACTTCGACATGTTGCTGACGACACACAGCGCGTGCACAC 1713
   : |||: ||| :|||: |||
Db 528 uSerCySAspLeuValGlnAsnCySThrIyTPheSerAlaSerPheSerHisSerMetAs 548
QY 1714 GCGCGCCCTGGCTGACGCTTACAGCTGACGCGCCCGAC-----GA 1755
   : |||: ||| :|||: |||
Db 548 p-----PhePheIeuIeuIySgylGlyProGlyValProMetValThrValHi 565
QY 1756 CGACCCCTTCACAGACAGCCCGCTTCTGGGCTAGCATGATGGAG---GCACGCCACTG 1812
   : |||: ||| :|||: |||
Db 565 saSnThrIthrAspIySlySlyMetPheAspLeuGluThrAsnGluIySValIySlySA 585
QY 1813 CCCCCCGGATTATGTTCTCTCCAGAGATCTTCATTTCCACACGCG---TGGATGTGGC 1869

```

```

Db 585 aIleAsnAspArgGlnMetProIySValGluTyrArgAspIleGluIleAspIyAs 605
QY 1870 GCTCTACGCGATGATTCACAGCCCGCTTGCACCGGAGGAAAGACGCCACCGT 1929
   : ||| :|||: |||: |||
Db 605 nLeuProMetGlnIleIeuIySProAlaIthrPheThrAspThrIthnIstYrProIleu 625
QY 1930 CCTCTTTGTATGAGAGCGCCCGGAGTGCAGCTGTGTAATACCTCTTCAAGGATCA 1989
   : ||| :|||: |||: |||
Db 625 uLeuValAlaAspGlyThrProGlySerGlnSerValAlaGluIySPhesGlu---ValSe 644
QY 1990 GTACTTCGCGCTCAACACACATCGGCTTCGAGTACCGGCTGATTCATTCAGTGCAG 2049
   : |||: ||| :|||: |||
Db 644 rTyrGluThrVal---MetValSerSerHisGlyAlaValAlaValIyScySAspGlyAr 663
QY 2050 GGGCTCTGTACGAGGAGCTTCCGTTTCCGAAGGCGCCGTGAAAACCAATGGCGCAGT 2109
   : |||: ||| :|||: |||
Db 663 gGlySerGlyPheGlnGlnIyThrIySLeuIeuHisGlyValAlaArgArgArgGlyLeuIe 683
QY 2110 GGAGATCGAGAGACAGGTGGAGGCGCTGCAAGTTCGTCGCGGAAAGATAGCTTCATCA 2169
   : ||| :|||: |||: |||
Db 683 uGluGluIySAspGlnMetGluAlaValAlaArgThrMetLeu---LysGluGlnIyTyrIleAs 702
QY 2170 CCTGAGCCGAGTTGCCATTCATGCGTGTCTACGAGGCGCTCTCTGCTCATGAGGCGT 2229
   : |||: ||| :|||: |||
Db 702 rArgThrArgValAlaValAlaPheGlyIySAspTyrGlyIyTyrIeuSerThrIyTyrIle 722
QY 2230 AATCCACAG-----CCCGAGTGTTCAGAGTGCATCGCGGCTGCCCGGT 2277
   : ||| :|||: |||: |||
Db 722 uProAlaIySgIyGluAsnGlnGlyIuThrPheThrCySgIySerAlaIeuSerProI 742
QY 2278 CACCGTGTGATGGCCCTACACACAGAGGTACAGCGCTACATGACGTCCCTGAGAA 2337
   : ||| :|||: |||: |||
Db 742 eThrAspPheIySLeuIyAlaSerAlaPheSerGluArgIyIeuIyLeuHisGlyLe 762
QY 2338 CAACACACAGCGCTATGAGGCGGCTTCCGCGCCCTGCACGTGAGAAAGGCCCAATGA 2397
   : ||| :|||: |||: |||
Db 762 uAspAsnArgAlaItyrIuMetThrIySValAlaHisArgValSerAlaIeu-----Gl 780
QY 2398 GCGCCACCGCTTGTCTTATCTCCACAGCGCTCTCTGAGAGAAACGTGCATTTTCCAC 2457
   : ||| :|||: |||: |||
Db 780 uGluGlnGlnPheIeuIleIeuIleHisProThrAlaAspGluIySHisIleHisIsth 800
QY 2458 AACTCTCTGTCTCCCACTGATCCGACAGGAGAAACCTTACAGCTCCAGATCTACC 2517
   : ||| :|||: |||: |||
Db 800 rAlaGluIleuIleThrGlnIleuIleArgGlyIySAlaAsnIySerIeuGlnIleTyrPr 820
QY 2518 CAACGAGAGACAGATATGCTGCTGCCCGCGAGTCCGGGAGGACATATGACATGCTGCT 2577
   : |||: ||| :|||: |||
Db 820 oAspGlnSerHisIyThrPheThrSerSerIeuIySglnHisIleuIyArgSerIleI 840
QY 2578 GCACCTTCTACAGGAA 2593
   : |||: ||| :|||: |||
Db 840 eaSnPhePheValGlu 845

```

RESULT 7
 DPP6_RAT
 ID DPP6_RAT STANDARD: PRT; 859 AA.
 AC P46101;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV like protein (Dipeptidyl aminopeptidase-
 DE related protein) (Dipeptidylpeptidase VI) (DPPX).
 GN DPP6.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;


```

Db 468 AsnAspAsn-----IleGln-----SerIleThrSerGlyAspIrr 479
QY 1493 GAGSTTTGGGAGGAGCGCTCCAAAGTCTGGGTCAATGAGGAGACCACTGGTGTAC 1552
Db 480 AspValThr-----GluIleuThrTyrAspGluLysArgAsnLysLeuTyr 495
QY 1553 TTCGAGGAGCAACGAGCAGCCGCTGAGACCACTCTTATGCGTGCATATGAGCG 1612
Db 496 PheLeuSerThrGluAspLeuProArgArgArgHisLeuTyrSerAlaAsn----- 512
QY 1613 GCCGCGAGATCGTAGCGCTTCCACAGCCCGCTTCCCATAGCTGC----- 1660
Db 513 -----ThrValAspAspPheAsnArgGlnCysLeuSerCysAsp 525
QY 1661 -----TTCATGAGCCGAGCACTTGCATATGCTGTC 1690
Db 526 LeuValGluAsnGlyTyrThrValSerAlaSerPheSerHisAsnMetAspPheLeu 545
QY 1691 AGCCACTACAGCAGCTGAGCAGCCGCC-----TGCCTGCAC----- 1729
Db 546 LeuLysCysGluGlyProGlyValAlaProThrValThrValHisAsnThrThrAspLysArg 565
QY 1730 ---GTCACAGCTGAGGAGCGCCGAGCAGCAGCCCTGCACAGCAGCCCGCTTCAG 1786
Db 566 ArgMetPheAspLeuGluAlaAsnGlu-----GluValGlnLysAlaIleTyr 581
QY 1787 GCTAGCATGATGAGGAGCAGCAGCTGCCCGCATTTATGTTCTCCAGACATCTTCAT 1846
Db 582 AspArgGlnMetProLysIle-----GluTyrArgLysIleGluVal----- 595
QY 1847 TTCACAGCGCGCTCGATGTCGCTTACGAGCATGATCAGTACAGCCCGCCTTCAG 1906
Db 596 -----GluAspTyrSerLeuProMetGlnIleLeuLysProAlaThrPheThr 611
QY 1907 CCAGGAGAGAGCAGCAGCCCGCTCTTTATATAGGAGCGCCCGCAGCTGAGTGTG 1966
Db 612 AspThrAlaHisTyrProLeuLeuValAlaAspGlyTyrProGlySerGlnSerVal 631
QY 1967 AATATACCTCTTCAAAAGCATCACTGCGGCTCAACACA-----CTGGCTCCTGT 2020
Db 632 SerIuArgPheGlu-----ValThrTrpGluThrValLeuValSerSerHis 647
QY 2021 GGCTAGCGCGCTGTGATGATGAGGAGCGGCGCTCTGCAGCGGCTTCGAGTCAA 2080
Db 648 GlyAlaValValValLysCysAspLysArgLysGlyPheGlnGlyThrLysLeuLeu 667
QY 2081 GGGGCGCTGAAAACCAATAGGCGGAGTGAAGTGAAGCAGCAGTGGAGCGCTGCAG 2140
Db 668 HisGluValArgArgArgLeuGlyPheLeuGlnGluLysAspGlnMetGluAlaValArg 687
QY 2141 TTCGTGGCGGAGAGTATGCTTATGACCTGAGCCGAGTGGCATGCGTGGTCC 2200
Db 688 ThrMetLeu---LysGlnGlnTyrIleAspLysThrArgValAlaValPheGlyLysAsp 706
QY 2201 TACGAGGCGCTCTCTGCTGCTGAGGCTAATTCACAG-----CCCAGGTG 2248
Db 707 TyrGlyGlyTyrLeuSerThrTyrIleLeuProAlaLysGlyGluAsnGlnGlyInThr 726
QY 2249 TTCAGAGTGGCCATCCGCGGTGCCCGGTGCACCTGTGATGGCTTACAGCAGGATAC 2308
Db 727 PheThrCysGlySerAlaLeuSerProIleThrAspPheLysLeuTyrAlaSerAlaPhe 746
QY 2309 ACTAGAGCTTACATGAGAGCTCCCTGAGAACACAGCAGCGCTATGAGCGGCTTCGCTG 2368
Db 747 SerGluArgTyrLeuGlyLeuHisGlyLeuAspAsnArgAlaTyrGlnMetThrLysLeu 766
QY 2369 GCGCTGACGCTGAGAGAGCTCCCAATGAGCCCAACGCTTATCTCCACAGCGCTTC 2428
Db 767 AlaHisArgValSerAlaLeu-----GluAspGlnGlnPheLeuIleHisAlaThr 784
QY 2429 CTGAGCAGAAAAGTGCACCTTTTCCACACAACAATCTCTGCTCCCACTGACGAGCA 2488
Db 785 AlaAspGluLysIleHisPheGlnHisThrAlaGluLeuIleThrGlnLeuIleLysGly 804

```

```

QY 2489 GGGAACTTACAGCTCCAGATCTACCCGAGAGACAGACAGTATCGCTGCCCGAG 2548
Db 805 LysAlaAsnTyrSerLeuGlnIleTyrProAspGluSerHisTyrPheHisSerValAla 824
QY 2549 TCGGGCGAGCAGTATGAGTCACTGCTGCACCTTTCACAGGAA 2593
Db 825 LeuLysGlnHisLeuTyrArgSerIleIleGlyPhePheValGlu 839

RESULT 8
ID DP66_BOVIN STANDARD; PRT; 863 AA.
AC P42659;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV like protein (dipeptidyl aminopeptidase-
DE related protein) (Dipeptidylpeptidase VI) (DP6X).
GN DP6.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX Bovidae; Bovinae; Bos.
NX NCBI_Taxid=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=92108018; PubMed=1729689;
RA Wada K., Yokotani N., Hunter C., Doi K., Wentholt R.J., Shimasaki S.;
RT "Differential expression of two distinct forms of mRNA encoding
RT members of a dipeptidyl aminopeptidase family";
RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: DP6X-L (SHOWN HERE) AND DP6X-S;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DP6X-L
CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DP6X-S IS FOUND IN
CC BRAIN, KIDNEY, OVARY AND TESTIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M76428; AAC41622.1; -.
DR EMBL; M76429; AAC41623.1; -.
DR MEROPS; S09_973; -.
DR InterPro; IPR002469; DP6X N-term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00326; Peptidase_S9; 1.
DR Pfam; PF00930; DP6X N-term; 1.
KW Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
FT DOMAIN 1 93
FT TRANSMEM 94 114
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 115 863
FT CARBOHYD 171 171
FT CARBOHYD 402 402
FT CARBOHYD 469 469
FT CARBOHYD 533 533
FT CARBOHYD 564 564
FT CARBOHYD 811 811
FT VARSPPLIC 1 79
FT PLGADPAAPARRENGGGGAGGRFRFYQARSDDDD ->

```



```

Db 775 SerAlaLeu-----GlUGlYgInGlnPheLeuValIleHisAlaThrAlaAspGluLys 792
QY 2441 GTGACGCTTTCCACACAACTTCCTGCTCCCAACGATGCCAGCGAGGAACCTTAC 2500
Db 793 IleHisPheGlnHisThrAlaGlnLeuIleThrGlnLeuIleGlyLysAlaAsnIyr 812
QY 2501 CAGCTCCAGATCTACCCCAAGACAGACAGTATTGCTGCTGCCGAGTCCGCGAGCAC 2560
Db 813 SerLeuGlnIleIyrTrpProAspGlnSerHisIyrPheSerAlaAlaLeuGlnGlnHis 832
QY 2561 TATGAACTGACGTTGCTGCTACCTTCTACAGGAA 2593
Db 833 LeuHisIyrSerIleLeuGlyPhePheValGlu 843
RESULT 9
SEPR_MOUSE STANDARD; PRT; 761 AA.
ID SEPR_MOUSE
AC P97321;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
  membrane serine protease).
GN FAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97284459; PubMed=9139873;
RA Niedermeyer J., Scanlan M.J., Garin-Chesa P., Daiber C., Flebig H.H.,
  Old L.J., Rettig W.J., Schnapp A.;
RT "Mouse fibroblast activation protein: molecular cloning, alternative
  splicing and expression in the reactive stroma of epithelial
  cancer."
RT Int. J. Cancer 71:383-389(1997).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Breast;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May have a role in tissue remodeling during development
  and wound healing, and contribute to invasiveness in malignant
  cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
  type IV collagen, but not native type I or type IV collagen. Does
  not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
  inactive (by similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms: 1 (shown here), 2 and 3; are
  produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Detected in fibroblasts, in placenta, uterus,
  embryos from day 7-19 and in new-born mice (P1).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y10007; CAAT7116.1; -.
DR EMBL: BC019190; AAH19190.1; -.
DR MEROPS: S09.007; -.
DR MGD: MGI:109608; FAP.
DR InterPro: IPR002469; DPPIV_N_term.
DR InterPro: IPR001375; Peptidase_S9.

```

```

DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR000379; Ser_estr_site.
DR Pfam: PF003326; Peptidase_S9; 1.
DR Pfam: PF00930; DPPIV_N_term; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
KW Hydrolase; Protease; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Alternative splicing.
FT DOMAIN 1 4 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 5 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
  (POTENTIAL).
FT DOMAIN 26 761 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 624 624 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 702 702 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 734 734 CHANGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 31 35 MISSING (IN ISOFORM 2).
FT VARSPPLIC 31 63 MISSING (IN ISOFORM 3).
FT CONFLICT 737 737 S->L (IN REF. 2).
SQ SEQUENCE 761 AA; 87944 MW; 9174C3AEDA213B25 CRC64;

```

Alignment Scores:

```

Pred. No.: 5.41e-15 Length: 761
Score: 438.50 Matches: 183
Percent Similarity: 37.45% Conservative: 126
Best local Similarity: 22.18% Mismatches: 287
Query Match: 9.14% Indels: 229
DB: Gaps: 35

```

US-09-976-674-4 (1-2617) x SEPR_MOUSE (1-761)

```

QY 362 CATGAGGCTCTCTCTCGGAGAGAGAGCTGTCGAGGAGCGGAACGCTGGGGTCTTC 421
Db 79 TyrAsnIleGlnThrArgGlnSerIyrIleIleLeuSerAsnSerThrmethylserval 98
QY 422 GGCATCACCTCTCTACGACTTC-----CAC 445
Db 99 AsnAlaThrAspIyrGlyLeuSerProAspArgInPheValIyrLeuGlnSerAspIyr 118
QY 446 AGCGAGATGGCTCTCTCTCTTCACGCCACCAACCTCTTCACTGTCGCGAGCG 505
Db 119 SerIyrLeuTrpArgIyrSerIyrThrAlaThrIyrIleIyrAspLeuGlnAsnGly 138
QY 506 GCGAAGAACGGCTTCATCGTGTCCCTATGAACCGGTGAATCAAGCAACGAGCTCA 565
Db 139 -----GluPheValArgGlyIyrGlnLeuProArgProIleGlnIyrLeuGlyStrp 155
QY 566 GGGCCCGGATGAGACCCCAAAATCTGC-----592
Db 156 SerPro--ValGlySerIyrLeuAlaIyrValIyrGlnAsnAsnIleIyrLeuGln 174
QY 593 ---CCTGCCAGACCTGCTCTCTCTTCATCAATACAGCGAGCTGTGGTGGCCAAC 649
Db 175 ArgProGlyAspProPhe-----GlnIle 183
QY 650 ATCGACAGCGAGGAGCGGCGGCTGACCTTCTGCACCAAGTTTATCCAAATGCTCTG 709
Db 184 ThrIyrThrGlyArgGlnAsnArgIle-----192
QY 710 GATGACCCCAAGTGTGCGGTGTGGCCACTTCTCATACAGAGAGAGCTTGCAGCGCTTC 769
Db 193 -----PheAsnGlyIleProAspIyrValIyrGlnGlnGlnMet---LeuAla 207
QY 770 ACTGGGTAC-----TGGTGGTGCGCCACAGCTCTCTGGGAAGGTTCAGAGGGCTCTAG 823
Db 208 ThrIyrIyrAlaLeuIyrTrpSerProAspGlyLysPhe-----220
QY 824 ACGTGCAGATCTCTATAGAGAGAGTCAATGATGATCCGAGTGGAGGTGAT---874

```

Db 221 -----LeuAlaTyrValGluPheAsnAspSerAspIleProIleIleAlaTyrSer 237
 QY 875 -----CAGCTGCCCTCTCGCTAGAGAAAGAGAGAGACTCGTAT 919
 Db 238 TyrTyrGlyAspGlyGlnTyrPro-----ArgThrIleAsnIle 250
 QY 920 CGGTACCCAGAGACAGACAGAAATCCCAAGATGCCCTTGAAATGCGTAGCTCCAG 979
 Db 251 ProTyrProIlyAsnGlyAlaIlyAsnProValValArgVal----- 264
 QY 980 ACTGACAGCCAGGCGAGATGCTGCACCCAGAGAAAGAGAGCTGTCAGCCCTTACG 1039
 Db 265 -----PheIleValAspThrThrTyrProHisValIleGlyProMetGlu 279
 QY 1040 TCGGTGTCCGAGAGTACATGATACAGCGCGGTGAGCCGGATGGCAATAC 1099
 Db 280 -----ValProValIleProGluMetIleAla-----SerSerAspTyrTyrPhe 293
 QY 1100 GCCTGGGCGCATGTCCTGAGACCGGCCAG-----CAGTGGCTCCAGCTGCTCTC 1150
 Db 294 SerTyrPheThrTyrValSerSerGluArgValCysLeuGlnTyrLeuIlyAsnArgValGln 313
 QY 1151 CTCGCCCGGCGCTTCATCCGACGACAGAGATAGAGAGCGGCTAGCCTTGCC 1210
 Db 314 AsnValSerValLeuSerTleCysAspPheArgGluAspTrpHis-----Ala 329
 QY 1211 AGAGCTGCCCGAGAGATGTCAGCCGCTATGCTGACGAGAGAGTACCAACGCTGG 1270
 Db 330 TyrPheIlyAsnProIlyAsn-----GlnGluHisVal-----GluIlySerArgThrGlyTyr 346
 QY 1271 ATCATTTTATGACATCTTATCCCTTCCCGCAATCAGAGGAGAGAGAGAGCTGTC 1330
 Db 347 AlaGlyGly-----PhePheValSerThrProAlaPheSerGlnAspAlaThrSer 363
 QY 1331 TTTTCGCGCCCAATGATGACAGACCGGCTTCGCATTTGTACAAGTACCGCGCTT 1390
 Db 364 TyrTyrIlyAsnIlePheSerAspIlyAspGlyTyrIlyHisIleHisTyrIle----- 380
 QY 1391 TTAAATCCAGGGCTACGATGAGTGCAGCCCTTCAGCCCGGGAGAGATTAATTAAG 1450
 Db 381 -----Lys 381
 QY 1451 TGCCCATTAAGAGAGATGCTCTGACAGCGGTGAATGGAGTTTGGCCAGGCAC 1510
 Db 382 AspThrValIleAsnAlaIleGlnIleThrSerGlyIlyStrpGluAlaIle----- 398
 QY 1511 GGCTCCAAAGTGGTCATGAGAGACCAAGCTGGTAC-----TTCAG 1558
 Db 399 -----TyrIlePheArgValThrGlnAspSerIleuPheTyrSerAsnGluPheGlu 416
 QY 1559 GGCACCAAGGACACCGCGCTGGAGCACCACTGACGTGCTATGAGCGCGCGC 1618
 Db 417 Gly-----TyrProGlyArgArgAsnIleTyrArgIleSerIleGlyAsnSerPro 433
 QY 1619 GAATGCTAGGCTCCACCGCGCGCTGTCSCATAGCTGCTCCAG----- 1666
 Db 434 ProSerIlyAsnCysValThr-----CysHisIleuArgGlyGluArg 447
 QY 1667 -----AGCCGAACTTCGACATGTCGTGACAGCACTGACAGAGAGCAGC 1714
 Db 448 CysGlnTyrTyrThrIleAspPheSerTyrIlyAlaIlyStrpTyrIlyAlaIleValCysTyr 467
 QY 1715 CGGCGCTGCGTACGCTCTCAACCTG--AGCGCGCCGACGACGACCGCTCCAC-- 1768
 Db 468 GlyProGlyLeuProIleSerThrLeuHisAspGlyArgThrGlnGluIleVal 487
 QY 1769 -----AAGCAGCCCGCG----- 1780
 Db 488 LeuGluGluAsnIlyGluLeuGluAsnSerLeuArgAsnIleGlnIleuProIlyValGlu 507
 QY 1781 -----TTCGGGCTAGCATGATGAGGAGCGCAGC 1810
 Db 508 IleIlyLysLeuIlyAspGlyGlyLeuThrPheThrTyrIlyMetIle----- 523

QY 1811 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGCGCTGGATGCGG 1870
 Db 524 -----LeuProProGln----- 527
 QY 1871 CTCTAGGCGATGATCTACAAAGCCCGCTTGACAGCCAGGGAAGAGACCCAGCGTC 1930
 Db 528 -----PheAspArgSerIlyLysTyrLeuLeu 537
 QY 1931 CTCCTTATATGAGGCGCCCGACGTCAGCTGGTGAATACCTCTTC-----AAAGGC 1994
 Db 538 IleGlnValTyrGlyGlyProCysSerGlnSerValIlySerValPheAlaValAsnTyr 557
 QY 1985 ATCAAGTACTGGCGCTCAACACACTGCGCTCCGTGGCTACGCGCTGTTGATTTGAC 2044
 Db 558 IleThrTyrIleu-----AlaSerIlyGlnIlyIleValIleAlaIleValAsp 573
 QY 2045 GGCAGGGCTCTGTCAGCGAGGCTTGCTGCGAGGGCGCTGAAAACCAATGGGC 2104
 Db 574 GlyArgGlyThrAlaPheGlnGlyAspIlyAspPheIleHisAlaValTyrArgGlyLeuGly 593
 QY 2105 CAGGTGAGATCCAGGACCGAGTGGAGGCGCTGCAGTTGCGCGGAGAGTATGCGTTC 2164
 Db 594 ValTyrGluValGluAspGlnLeuThrAlaValArgIlySphIleGlu--MetGlyPhe 612
 QY 2165 ATCGACCTGAGCGGAGTGGCATTCAGTGCCTGCTACGAGGGGCTTCCTGCTCATG 2224
 Db 613 IleAspGluGluArgIleAlaIleTyrPheTyrSerTyrGlyGlyTyrValSerSerLeu 632
 QY 2225 GCGCTAATCCAAAGCCCGAGGCTTCAGAGTGGCCATCGCGGCTGCCGCTACCGCTC 2284
 Db 633 AlaLeuAlaSerGlyThrGlyIlePheIlyCysGlyIleAlaValAlaProValSer 652
 QY 2285 TGATGGCTACGACAGGCTACATGAGCGGTGACATGAGCGTTC-----GAGAAC 2338
 Db 653 TyrPheTyrTyrAlaSerIleTyrSerGluArgPheMetGlyLeuProThrIlyAspAsp 672
 QY 2339 AACGACGACGCTATGAGCGGCTCCGTGCGCTGACAGTGAAGAGTCCCAATGAG 2398
 Db 673 AsnLeuGlnHisTyrIlyAsnSerThrValMetAlaArgAlaGluTyrPheArgAsnVal 692
 QY 2399 CCCAACGCTTGCTTATTCCTCCAGCGCTCTCGACGCAAAACGTGCTTTTCCACACA 2458
 Db 693 Asp-----TyrLeuLeuIleHisGlyThrAlaAspAsnValHisPheGlnAsnSer 710
 QY 2459 AACTCTCGCTCCCAAGTATCGGACGAGCAAACTTACCACTGACATGATACCCC 2518
 Db 711 AlaGlnIleAlaIlyAlaIleValAsnAlaGlnValAspPheGlnAlaMetTyrTyrSer 730
 QY 2519 AACGAGACACAGTATTCGCTCGCCCGAGTCCGCGAGACATATGACATGCTGCTG 2578
 Db 731 AspIleAsnHisGlyIleSerSerGlyArgSerGlnAsnHisLeuTyrThrHisMetThr 750
 QY 2579 CACTTCTACAGAA 2593
 Db 751 HisPheLeuIlyGln 755
 RESULT 10
 DAP2_YEAST
 ID DAP2_YEAST STANDARD; PRT; 818 AA.
 AC P18962;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DAP B) (YSCV).
 GN DAP2 OR YHR028C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89174971; PubMed=2647766;

RA	Robertis J., Pohlig G., Rothman J.H., Stevens T.H.;
RT	"Structure, biosynthesis, and localization of dipeptidyl
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast
RT	vacuole."
RL	J. Cell Biol. 108:1363-1373(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	SPRAIN-S288C / AB972;
RC	MEDLINE=94378003; PubMed=8091229;
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA	Ku D., Favella A., Fulton L., Galtong S., Geisel C., Kirsten J.,
RA	Kudava T., Hillier L., Jier M., Johnston L., Langston Y.,
RA	Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA	Nhan M., Rifkin H., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA	Vaudin M.;
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT	viii."
RL	Science 265:2077-2082(1994).
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYSOSOME-LIKE
CC	VACUOLES.
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X15484; CAA3512.1; -;
DR	EMBL; U10399; AAB68879.1; -;
DR	PIR; A30107; A30107.
DR	PIR; S46780; S46780.
DR	MEROPS; S09_006; -;
DR	SGD; S0001070; DAP2.
DR	InterPro; IPR002469; DPTIV_N_term.
DR	InterPro; IPR001375; Peptidase_S9.
DR	InterPro; IPR002471; Prol_endopep_ser.
DR	InterPro; IPR000379; Ser_estr_sste.
DR	Pfam; PF003326; Peptidase_S9; 1.
DR	Pfam; PF00930; DPTIV_N_term; 1.
DR	PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
DR	HydroLase; Aminopeptidase; Dipeptidase; Serine protease;
KW	Transmembrane; Glycoprotein; Signal_anchor.
FT	DOMAIN 1 29
FT	TRANSMEM 30 45
FT	DOMAIN 46 818
FT	DOMAIN 679 679
FT	ACT_SITE 756 756
FT	ACT_SITE 789 789
FT	CARBOHYD 63 63
FT	CARBOHYD 79 79
FT	CARBOHYD 110 110
FT	CARBOHYD 139 139
FT	CARBOHYD 372 372
FT	CARBOHYD 392 392
FT	CARBOHYD 421 421
FT	CARBOHYD 738 738
FT	CONFLICT 83 83
FT	CONFLICT 125 125
FT	CONFLICT 182 188
FT	CONFLICT 200 200
FT	CONFLICT 366 375
FT	CONFLICT 808 818
SO	SEQUENCE 818 AA; 93404 MW; 318F450445375BD3 CRC64;
Alignment Scores:	
Pred. No.:	1.09e-14
Score:	432.50
Percent Similarity:	39.47%
Length: 818	
Matches: 164	
Conservative: 91	

Best Local Similarity:	25.39%	Mismatches:	262
Query Match:	9.02%	Indels:	129
DB:	1	Gaps:	28
US-09-976-674-4 (1-2617) x DAF2_YEAST (1-818)			
QY 728	GGTGTGGCCACACCTTCGTCATACAGGAAGAG---TTGACCGGTTCACTGGGACTGGTG	784	
Db 234	GIYLSProAspTrpValTYTGUGUGUGValPheGluAspLysAlaLaIaTrpTr	253	
QY 785	TGCCCCACACGCTCCCTGGAGAGGTCAGAGGCGCTCAAGACGCTCGAATCCTGATAG	844	
Db 254	SeTrpThrGluAspTrpY---	263	
QY 845	GAAGTCGATGAGTCCGAGGTGAGGTCAATTCAGTCGCTCCCTGCGCTAGAGAAAG	904	
Db 264	LysIleAspGluSerGluValGluPheIleIleProTYTValGluAspTrpLys	283	
QY 905	-----AAGACGACCTGTATCGGTACCCGAGACAGGACAGCAAGATCCCAAG	952	
Db 284	AspIleTYrProGluMetArgSerIleLysTYrProLysSerGlyThrProAsnProHis	303	
QY 953	ATTCGCTTGAACATCGCTGAGTTCGACAGACTGACAGCGCAGGCAAGATGCTTCGACAG	1012	
Db 304	AlaGluLeuTrpValTYrSerMetLysAspGly-----	314	
QY 1013	GAGAAAGAGCTGTGTGACGACCCCTTCAGCTGCTCCCTCCGAGGTGAG-----	1060	
Db 315	-----ThnSerPheHisProArgIleSerGlyLysLysLys	326	
QY 1061	-----TACATCGCCAGAGCGCGGTGGACCCGGGATGGCAATAGCCCTGGGCC	1108	
Db 327	AspGlySerLeuLeuIleThrGluValThrTrpValGluLysGluLysAn--ValLeuVal	345	
QY 1109	ATGTCCTCGAGCGCGGCCCGACAGAGTGGCTCCAGCTGCTGCTCCCTCCCGCGGCTGTTC	1168	
Db 346	LysThrThrAspArgSerSerAspIleLeuThrValPheLeuIleAspThr-----	362	
QY 1169	ATCCCGAGACAGAAATGAGAGACAGCGGCTAGCCTTCGACAGACTGTCCCGAGAAAT	1228	
Db 363	IleAlaLysThrSerAsn-----	368	
QY 1229	GTCCAGCGCGTATGGTGTTGTCAGAGAGGTACCAAC-----GTCTGATCAATGTTAT	1283	
Db 369	-----ValValArgAsnGluSerSerAsnGlyGlyTYrTrpIleThrHis	384	
QY 1283	GACATCTTCATCCCTTCCCGCAATCAGAGGAGAGAGAGAGAGCTGCTTTCGCGGCC	1342	
Db 385	AsnThrLeu-----PheIleProAla	391	
QY 1343	AATGATATGC-----AAGACCGGCTTCGCAATTTGACAAATCCACCGGCT	1390	
Db 392	AsnGluThrPheAspArgProHisAsnGlyTYr-----ValAspIle	405	
QY 1391	TTAAATATCCAGGCTACGATTTGAGTGAAGCCCTTC---AGCCCCGGGAGATCAATTT	1447	
Db 406	LeuProIleGlyGlyTYrAsnHisLeuAlaTYrPheGluAsnSerAsnSerHisTYr	425	
QY 1448	AAGTCCCCCATTAAGAAAGATGCTCTTACACAGCGGTAAGGAGAGGTGTTGGCGAGG	1507	
Db 426	Lys-----ThrLeuThrGluGlyLysTrpIleValVal-----	436	
QY 1508	CACGGCTCCAAATCTGGATCAATGAGAGAGACCAAGCTGGTACTTCCAGGGCACCAAG	1566	
Db 437	AsnGluProLeuAlaPheAspSerMetGluAsnArgLeu---TYrThrIleSerThrArg	455	
QY 1568	GACAGCGCGCTGGAGACACACACTTACGTGTGATCACTATGAGGCGCGGCGAGATGTA	1622	
Db 456	LysSerSerThrGluArgHisValTYrTYrIleAspLeuArgSerProAsnIleIle	475	
QY 1628	CGCCCTCACAGCGCCCGG-----TTCTCCCAATAGAGGCTCTGACGACAGAGAC	1675	
Db 476	GluValThrAspThrSerGluAspGlyValTYrAspValSerPheSerSerGlyArgArg	495	

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: U09278; AAB49652.1; -
 DR EMBL: U76833; AAC51668.1; -
 DR EMBL: AF007822; AAF21600.1; -
 DR MEROPS: S09.007; -
 DR Genew: HGNC:3590; FAP.
 DR MIM: 600403; -
 DR InterPro: IPR002469; DEPIV_N-term.
 DR InterPro: IPR001375; peptidase_S9.
 DR InterPro: IPR002471; ProL_endopep_ser.
 DR InterPro: IPR000373; Ser_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR Pfam: PF00930; DEPIV_N-term; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 KM Hydrolyase: Protease; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Alternative splicing.
 FT DOMAIN 1 4
 FT TRANSMEM 5 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 26 760
 FT ACT_SITE 624 624
 FT ACT_SITE 702 702
 FT ACT_SITE 734 734
 FT CARBOHYD 49 49
 FT CARBOHYD 92 92
 FT CARBOHYD 99 99
 FT CARBOHYD 314 314
 FT CARBOHYD 679 679
 FT VARSPIC 1 521
 FT CONFLICT 207 207
 FT CONFLICT 229 229
 FT CONFLICT 354 354
 FT SEQUENCE 760 AA; 87820 MM; AOD34B4801BE07EA CMC64;
 Alignment Scores:
 Pred. No.: 2.28e-14 Length: 760
 Score: 426.00 Matches: 168
 Percent Similarity: 38.85% Conservative: 109
 Best Local Similarity: 23.56% Mismatches: 256
 Query Match: 8.88% Indels: 180
 DB: Gaps: 31

US-09-976-674-4 (1-2617) x SEPR_HUMAN (1-760)

OY 593 CCTGCGACCTGCTTCTCTCTCATATAACAGCAGCAGCTGGGTGCCCAATC 652
 DB 176 ProLysAspProPhe----- 181
 OY 653 GAGACAGCGAGGCGCGCTGACCTTCCACACAGGTTTATCCATGCTCGAT 712
 DB 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnIlePhe 193
 OY 713 GACCCCAAGTGTGGGGTGGCCACTTGTCTATACAGAGAGACTTGCACGCTTCACT 772
 DB 194 Asn-----GlyIleProAspTyrPValTyrGluGluGluMet---LeuProThr 208
 OY 773 GGGTAC-----TGGTGTGGCCCAACAGCTCCGCGGAAGTTGACAGGGCCTCAAGAC 826
 DB 209 LysTyrAlaLeuTyrPheSerProAsnGlyLysPhe----- 220
 OY 827 CTGGCAATCCGTATGAGAGAGATGATCCGAGGTGAGGTGATTCACGTCCTCTCT 886
 DB 221 -----LeuAlaTyrAlaGluPheAsnAspLysAspIleProValIleAlaTyrSerTyr 238
 OY 887 CCTCGCGTAGAGAA--AGGAGACGAGACTGCTATCGGTACCCAGAGACAGCAGCAAG 943
 DB 239 TyrGlyAspGluGlnTyrProArgThrIleAsnIleProTyrProLysAlaGlyAlaLys 258

OY 944 AATCCCAAGATTGCTTGAATACTGCTGAGTTCACAGACTGACAGCCAGGCAATCGTC 1003
 DB 259 AsnProValAlaArgIle-----PheIleIle 267
 OY 1004 TCGACCCAGAGAAAGAGACTGTGCGACCCCTCAGCTGCTGTCCGAAGTG----- 1057
 DB 268 AspThrThrTyrProAlaTyrValGlyProGluGluValProValProAlaMetIleAla 287
 OY 1058 -----GAGTACATGCGACGCGCGGGGTGGACCCGGGATGCAAAATCCGTGGGCC 1108
 DB 288 SerSerAspTyrTyrPheSerThrPheThrValThrAspGluArgValCys----- 305
 OY 1109 ATGTTCCGAGACCGGCCCCAGACAGGCTCCGACGCTGCTCTCTCTCTCTCTCTCT 1168
 DB 306 -----LeuGlnIlePheLysArgValGlnAsnValSerValLeuSer 319
 OY 1169 ATC-----CCGAGCAGCAGAAAT----- 1186
 DB 320 IleCysAspPheArgGluAspThrPheThrThrPheAspCysProLysThrGlnGluHisIle 339
 OY 1187 GAGAGCAGCGGCTA-----GCTCTGCGCAGAGCTTCCCGAGAATGTCACGCGTAT 1240
 DB 340 GluGluSerArgThrGlyTyrPheValGlyPhePheValSerArgProValPheSerTyr 359
 OY 1241 GTGGTGTACGAGGAGGTGACCAACGCTGAGCATGTTCAAGACATCTTATCCCTTC 1300
 DB 359 ----- 359
 OY 1301 CCCCAATCAGAGGAGAGAGAGAGAGCTGCTTCTCCGCGCAATGAATGCAAGACCGGC 1360
 DB 360 -----AspAlaIleSerTyrTyrLysIlePheSerAspLysAspGly 373
 OY 1361 TTCTGCCATTGTACAAAGTACCGCCGCTTTAAATCCAGAGGCTCAGATTGAGTGA 1420
 DB 374 TyrLysHisIleHisTyrIle----- 380
 OY 1421 CCTCTACAGCCCGGGAGAGATGATTAAGTCCCATTAAGAGAGATTGCTCTGAC 1480
 DB 381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
 OY 1481 AGCGGTAGTGGAGGTTTGGCGAGCAGCGCTCCAGATGTCGTCGTCATGAGAGACC 1540
 DB 392 SerGlyLysTyrPheGluAlaIle-----AsnIlePheArgValThrGlnAsp 406
 OY 1541 AAGCTGTACTTCCAGGGCACC---AAGACACGCGCTGAGACACCACTCTAC--- 1594
 DB 407 SerLeuPheTyrSerSerAsnGluPheGluGluTyrProGlyArgArgAsnIleTyrArg 426
 OY 1595 -----GTGGTCAGCTATGAGCGCGCGCGAGATCGTACGCTCACCAGCGCGGCTTC 1648
 DB 427 IleSerIleGlySerTyrProSerLysCysVal----- 439
 OY 1649 TCCCATAGCTGCTCATG-----AGCCAGCAACTTCGACATG 1684
 DB 440 -----ThrCysHisLeuArgLysGluArgCysGlnTyrTyrThrAlaSerPheSerasp 457
 OY 1685 TTCGTACCCACTACAGCGGTGAGACGCGCGCTCGTCGTCACGTCATCAAGCTG--- 1741
 DB 458 TyrAlaLysTyrTyrAlaLeuValCysTyrGlyProGlyIleProIleSerThrLeuHis 477
 OY 1742 AGGCGCCCGCAGAGACCCCTGCAC-----AAGCAGCCCGCTTCTGGCTGACATG 1795
 DB 478 AspGlyArgThrAspGluGlnIleLysIleLeuGlnAsnLysGluLeuGluAsnAla 497
 OY 1796 ATGAGCAGCAGCAGCTGCCCGCGGATTAATGCTCCAGAGATTTCCATTTCCACACG 1855
 DB 498 LeuLysAsnIleGlnLeuProLysGlu-----GluIleLysLysLeuGluVal 513
 OY 1856 CGCTCGATGTGCGGCTTACGGCATCTTAAGCCCAAGCCCGCTTCCAGCCAGGAG 1915
 DB 514 AspGluIleThrLeuThrTyrLysMetIleLeuProGln---PheAspArgSerLys 532

```

QY 1916 AAGCACCACCCCTCTTGTATATGAGAGCCGCCAGGTGAGCTGGTGAATAACTCC 1975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 LysTrpLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1976 TTC-----AAGGATCACTACTGTGGGCTCAACACACTGCTCCCTGCTAGCC 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGluGlyMetVal 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2030 GTGGTGTGATGATGAGCGGGGCTCTGTCAGCGAGGCTTCGATCGAAGGCCCTG 2089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyAspLysLeuTyrAlaVal 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2090 AAAAACCAATGGCCAGGTGAGATGAGAGCAGGAGGAGGAGGAGGAGGAGGAGG 2149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 TyrArgLysLeuGlyValTyrGluValGluAspGlnIleThrAlaValArgLysPhe 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2150 GAGAAAGATGAGCTTATGAGAGCTGAGCGGAGTTCATCCATGGCTGCTTACGGG 2209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 Glu--MetGlyPheIleAspGluLysArgIleAlaIleTrpGlyTrpSerTyrGly 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2210 TTCCTCTGCTATGAGGCTATATCCAGAGCCAGGCTGTCAAGTGGCCATGCGGG 2269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 628 TyrAlaSerLeuLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2270 GCCCGGTCACCGTGTGATGCGCTACGACAGGAGTAACTGAGCGCTATGAGAG 2329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 648 AlaProValSerSerTrpGluTyrTyrAlaSerValTyrThrGluArgPheMetGly 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2330 CCT-----GAGAACACACAGCAGGCTATGAGCGGCTTCGCTGCGCTCAGCTG 2383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 ProThrLysAspAsnLeuGlnIleTyrLysAsnSerTrpValMetAlaArgAlaVal 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2384 AAGCTGCCAATGAGCCCAACCGCTTATCCATGAGCTTCCTGCTGAGAAACGTG 2443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAsnVal 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2444 CACTTTTCCACACAACTTCCTGCTCCCACTGATCCGAGAGGAAACCTTACAG 2503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 HisPheGlnAsnSerAlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGln 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2504 CTCGAGATCAACCCCAACGAGACAGACATTTGCTGCTCCCGAGCTGGGCGC----- 2554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 AlaMetTrpLysSerAspGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2555 GAGCCTATGAGTACAGTGTGCTGCTGCTTCTACAGAA 2593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 AsnHisLeuTyrThrHisMetThrHisPheLeuLysGln 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RL (In) Gething M.-J., Novick P., Stevens T.H., Rothblatt J. (eds.):
RL Guidebook to the yeast secretory pathway, pp.1-1, Oxford University
RL Press, Oxford (1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=5286c / FY1679;
RX MEDLINE=96437977; PubMed=8840505;
RA Galissou F., Dujon B.;
RT "Sequence and analysis of a 33 kb fragment from the right arm of
RT chromosome XV of the yeast Saccharomyces cerevisiae.";
RL Yeast 12:877-885(1996).
CC -1- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC MATURATION OF THE
CC ALPHA-FACTOR PRECURSOR.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, LYSOSOME-LIKE
CC VACUOLES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L21944; AAA35119.1; -
CC EMBL: U08230; AAA17897.1; -
CC EMBL: X92441; CAA63182.1; -
CC EMBL: Z75127; CAA99437.1; -
CC PIR: A49737; A49737.
CC PIR: S45451; S45451.
CC MEROPS: S09.005; -.
CC SGD: S0005745; STE13.
CC InterPro: IPR002469; DDPV N.term.
CC InterPro: IPR001375; Peptidase_S9.
CC InterPro: IPR002471; ProL_endopep_ser.
CC InterPro: IPR000379; Ser_estrs_site.
CC Pfam: PF00326; Peptidase_S9; 1.
CC Pfam: PF00930; DDPV N.term; 1.
CC PROSITE: PS00708; PRO_ENDOPEP_SER; FALSE NEG.
CC KW Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
CC Transmembrane; Glycoprotein; Signal anchor; Pheromone response.
CC FT DOMAIN 1 119
CC TRANSMEM 120 140
CC FT DOMAIN 141 931
CC FT ACT_SITE 785 785
CC FT ACT_SITE 863 863
CC FT ACT_SITE 896 896
CC FT CARBOHYD 377 377
CC FT CARBOHYD 814 814
CC SEQUENCE 931 AA; 107200 MW; 81AE70094093C023 CRC64;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
6.68e-14	931	226	226	125	282	42
Score:	417.00					
Percent Similarity:	35.78%					
Best Local Similarity:	23.04%					
Query Match:	8.70%					

```

US-09-976-674-4 (1-2617) x DAPI_YEAST (1-931)
QY 33 CGGCGGACGAGGAGGAGCA-GCCGACAGATGAGCCGCGCGCTTCAGGTGACAG 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 ArgProThrGluAlaThrIleAspValThrAspValProGlnThrProPheLeuGlnGlu 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 92 AAGCAGCTGTGGAGGAGGCTCGAGGATCATCCAGGAGCGAGCGCAAGTACTCGGGCTC 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 GlnTyrSerMetArgProArgArg-----
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 152 ATTGTCAACAGGCGCCGACGATTCAGTTTGTGCAAGAAGAGATGACTGCGGCC 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 -----GluSerPheGlnPhe-----AsnAspIleGluAsnGlnHis 94
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

Db 707 -----ThrPheThrThyLysSerSerLeuAlaPheGluGlnAlaVala 720
QY 2038 GATTGACGGC-----AGGGGGCTCCGTGACGAGG 2067
Db 720 ValSerGlyLeuAspValIleValLeuGlnIleGluProArgGlyThrIleGlyLysG1 740
QY 2068 GCATTCGGTTTCGAGAGGGCCCTGAAAAACCAATAGGGCAGGTAGATCGAGACCAAGT 2127
Db 740 YTPSerPheArgSerThrAlaArgIleGlyLysLeuGlyTyrThrLeuProArgSpleth 760
QY 2128 GGAG---GGCCTCAGTCTCGCCGCGAGAAAGTATGGCTTATGACCTGACCGAGTTC 2184
Db 760 rgluValThrLysLysPheIleGlnArgAsnSerGlnHisIleAspIleuSerLysIleAl 780
QY 2185 CATTCCATGGCTGGTCCACGAGGGGCTTCCTGGCTCATGCGGCTA---ATCCACAAAGCC 2241
Db 780 aileTrrpGlyTrrSerTrrGlyTrrPheTrrSerLeuLysThrValIleuAspAsnG1 800
QY 2242 CCAGGTCTTCAGAGTGCATCCGCGGGTCCCGGTCCAGTCTGATGGCTTACGACAC 2301
Db 800 yaAPThrPheLysTyrIleAlaValAlaIaProValThrAsnThrPheLysTrrAspse 820
QY 2302 AGGCTACACTGACGCGCTTACATGACGTCCCTGAGAACACACAGCAGCGCTATGAGCGCG 2361
Db 820 rValTyrThrGluArgTyrMetAsnGlnProSerGluAsnHisGluIleGlyTrrPheGluVa 840
QY 2362 TTCGCGTGGCCCTGCAGCGTGAAGAAGTGGCCCAATGACCCCAAGCTTGGTATTCCTCCA 2421
Db 840 lserThrIleGlnAsnThelysSerPhe---GluSerLeuLysArgLeuPheIleValH1 859
QY 2422 CGGCTTCCTGAGCAAGAACTGTCATTTTTCACACAAACTCTCTCTCCCAACTGAT 2481
Db 859 sgLyThrPheAspAspAsnValHisIleGlnAsnThrPheArgLeuValAspGlnLeuAs 879
QY 2482 CCGAGCAGGG---AAACCTTACACGCTCCAGATCTACCCCAAGCAGAGACAGTATTCG 2538
Db 879 nleuLeuGlyLeuThrAsnTyrAspMetHisIlePheProAspSerAspHisSerIleAr 899
QY 2539 C 2539
Db 899 g 899

RESULT 13
AMNH_YEAST STANDARD; PRT: 1367 AA.
AC P08640: P08068:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE glucanase 51/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DN glucosidase) (1,4-alpha-D-glucan glucosylhydrolase).
GN STAL OR STAZ OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP STRAIN=S288c / AB972;
RC Barril B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churher C.M., Connor R., Copsey T., Hunt S., Devlin K., Fraser A.,
RA Gentes S., Hamlyn N., Horsnell T.S., Dear S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RP MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
STAL.";

```

```

RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STX2 and SGA genes
from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MDC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 1367
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BBD6D1AA9D CRC64;

Alignment Scores:
Pred. No.: 5,38e-11 Length: 1367
Score: 359.50 Matches: 216
Percent Similarity: 36.08% Conservative: 104
Best Local Similarity: 24.35% Mismatches: 380
Query Match: 7.50% Indels: 187
DB: Gaps: 30

US-09-976-674-4 (1-2617) x AMNH_YEAST (1-1367)
QY 9 CCATGGCCACACCGGAGCCCAACGCGCGAGCGAGCGCCGACAGATGACC 68
Db 272 ProThrProThrThrThrThrThrSerCysThrLysGluLysProThrProThrProH1SH1Asp 291
QY 69 CGGCGCGCGCGCTTCACAGTGCAGACACTCGTGGAGCGGCTCCGAGATATCCAGC 128
Db 292 ThrThrProCysThrLysLysThrThrThrSerLysThrCysThrLysLysThrThr 311
QY 129 GCA-----GGCGCAAGTACACGGGCGCCATGTCACAAAGCGCCCGCCAGC 173
Db 312 ThrProValProThrProSerSerThrThrThrGluSerSerAlaProValProThr 331
QY 174 ACTTCCAGTTTGTGCAAGAAAGAGATGATGTCGGCCCACTCCACCGCGCTTACTACAC 233
Db 332 ProSerSerSerThrThrThrGluSerSerSer-----AlaProValThrSerSerThr 348
QY 234 TGGGATGCCATATGCGACCGCGAGACATCCCTCTACTCTG-----AGA 281
Db 349 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSer 368
QY 282 TTCACAAAGAGTCCGGAAGAGAGGCTGCTGCTCCGCTCT-----GGAAGCAGATGC 335
Db 369 AlaProValThrSerSerThrThrThrGluSerSerSerAlaProValThrSerSerThrThr 388

```

OY	336	TGATCATTTCCAGGCCACGCCCCAC	---ATGGGCTTACTCTCGGAGAGAGACTGC	392
Db	389	glUserSerSeriAlaProValProThrProSerSerSerThrThrGlUserSerSeriAla		408
OY	393	TGAGGACCGGAAACGGCTGGGGGCTCTTGCGGAAACACCTCTTCACGACTTCCACAGCGANA		452
Db	409	ProValThrSerSerThrThrGlUserSerSeriAlaProValThrSerSerThr		426
OY	453	GTGGCCTTCCTCTTCACAGGCCACGACAGACCCCTTCCTCCGACGAGCGCGANA		512
Db	427	---ThrGlUserSerSeriAlaProValThrSerSerThrThrGlUserSerSeriAlaPro		445
OY	513	ACGGCTTCATGCTGCCCTATGCAAAACCGCTGGAATCAAGACCAAGCTCAGGGCCC		572
Db	446	ValThrSer	---SerThrThrGlUserSerSeriAlaPro	457
OY	573	GGATGGACCCCAAAATCGCCCTCCGACCCCTGCCCTTCTCTCT	---TCATCAATA	626
Db	458	ValProThrProSerSerSerThrThrGlUserSerSeriAlaProValThrSerSerThr		477
OY	627	ACAGGACCTCTGGGTGGCCCAATCGACATCGACAGACGCGCGCGCTGACCTCTGCC		686
Db	478	ThrGlUserSerSeriAlaProValProThrProSerSerSerThrThrGlUserSerSer		497
OY	687	ACCAAGCTTATTCCAATGCTCCGTGATGACCCCAAGTCTCGGGGTGTGGCCACTTGCTGA		746
Db	498	AlaProValThrSerSerThrThrGlUserSerSeriAlaProValProThrProSer		516
OY	747	TACAGAGAAGATTTGACCCGCTTCACTGCTGACTGTGTGTCGCCACAGCCTCTCGGGAAG		806
Db	517	---SerSerThrThrGlUserSerSeriAlaProValProThrProSerSerSer		533
OY	807	GTTCCAGAGGCGCTCCAGACGCGTGGCAATCTCTGATAGAGAAGTCATGATGCCAGATGG		866
Db	534	ThrThrGlUserSer		538
OY	867	AGGTCAATTCACGTCCTCCCTCTCCGCTAGAGAAAGAGAGACGACTCGTATCGGTACC		926
Db	539	---SerAlaProVal	---ThrSerSerThrThr	547
OY	927	CCAGACAGCAGCGACGANAAGATCCCAAGATTGCTTGAATCTGCTGATGCCAGACTGACA		986
Db	548	glUserSerSeriAlaProValPro		555
OY	987	GCCAGGGCAAGATGCTGTCGACCCAGGAGAGAAGAGCTGTGCGACGCCCTTCACACTGCTGT		1044
Db	556	---ThrProSerSerSerThr		561
OY	1047	TCCCGAAGGTGGATGCATCGACGAGGCGCGGTGAGACCGCGGATGCAAAATACGCTGGG		1106
Db	562	ThrGlUserSerSerThrProValThrSerSerThrThrGlUserSerSeriAlaProVal		581
OY	1107	CCATGTTCTCGAGCCGCGCCAGCAGTGGCTCAGTCTCTCTCTCC	---CCC	1157
Db	582	ProThr	---ProSerSerSerThrThrGlUserSerSeriAlaProValPro	597
OY	1158	CGGCGCTGTTCATCCGACGACGACAGAGATAGAGAGAGAGCGGCTAGCCTTCCACAGACTG		1211
Db	598	ThrProSerSerSerThrThrGl	---SerSerSeriAlaProValPro	612
OY	1218	TCCCCAGGAATGTCACCGCTATGTGTGACGAGGAGGTCACCAACGTCGTGATCAATG		1277
Db	613	ThrProSerSerSerThrThrGl	---SerSerSeriAlaProValThrSerSerThr	630
OY	1278	TTTCATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGACGAGACTCTCTTCTCC		1337
Db	631	ThrGlUserSerSeriAlaProValProThrProSerSerSerThrThrGlUserSerSer		650
OY	1338	GGCGCATGATGCAACACCGGCTTCTGCATTTGTACAAAGTCACCGCGCTTTAAAT		1397
Db	651	AlaProValProThrProSerSerSerThrThrGlUserSerSeriAlaProValProThr		670

QY	1398	CCGACGGCTACGATTGGAGAGGAGCCCTTCAGCCCGCGGGAAGATGAATTAAGTCCCA	1457
Db	671	ProSerSerSerThrThrGluSerSerSerAla	681
QY	1458	TTAAGAAGAGATTGCTCTGACACGAGCGGTAAATGGAGAGTTTGGGAGACGAGCTCCA	1517
Db	682	-----ProValThrSerSerThrThrGluSerSerSerAlaPro	694
QY	1518	AGATCTGGGTCATGAGGAGACCAAGCTGGTGTACTTCCAGGGCACCAAGACACGCCGC	1577
		:::	
Db	695	ValThrSerSerThrThrGluSerSerSerAlaProValProThrProSerSer	712
QY	1578	TGGAGACACCACTCTAGCTGGTACGTATGAGGGCGCGCCGACAGTCGACGCTCACCA	1637
Db	713	--SerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu	731
QY	1638	CGCCGGCTTCTCCATGAGTGGCTGCACATGAGCCAGAACTTCGACATGTTCCGACCCACT	1697
		:::	
Db	732	SerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAlaPro	751
QY	1698	ACAGCAGCGTGAAGCAGCCGCCCTGCGTGCACGCTACAGCTGAGCGGCCCGCCGACGAG	1757
		:::	
Db	752	ValThr-----SerSerThrThrGluSerSerSerAlaProValPro	765
QY	1758	ACCCCTTCGACACAGCAGCCCGCTTCTGGGCTAGCATATGAGAGCAGCAGCTCCGCC	1817
Db	766	ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSer	784
QY	1818	CGGATTATGTTCTTCGAGAGATCTTCGATTTCACACGCGCTCGGATGTGGCGCTCTACG	1877
Db	785	-----SerThrThr	787
QY	1878	GCATGATCTACAAAGCCCGACCTTCGAGCCAGGAGAACCCGACCGCTCTTGG	1937
Db	788	GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerValAla	807
QY	1938	TATATGAGAGCGCCCGACGCTGACGTGGTGAATACTCTCTCAAGAGCATCAAGTCTGC	1997
		:::	
Db	808	ProValProThrProSerSerSerSerSerAlaProSerSerAlaProSerSerThrPro	827
QY	1998	GGCTCAACACATGGCT-----CCCTGGGCTACGCCGTGGTGTGATTGACG	2045
Db	828	PheSerSerSerThrGluSerSerSerValProValProThrProSerSerSerThrThr	847
QY	2046	-----CGAGGGGCTCTGTGCAGCGAGGGCTC-----GGTTCGAAGGGCGCCGGA	2093
		:::	
Db	848	GluSerSerSerAlaProValSerSerSerThrThrGluSerSerValAlaProValPro	867
QY	2094	ACCAAAATGGGCGACGTGAGATCGAGAGACAGGTGAGGGCTCGCATGTTGTCGTCGAGA	2153
Db	868	Thr-----ProSer	870
QY	2154	AGTATGGCTTCATCGACCTTACGCCGAGTTCCTCATGCTGGTGCCTACGAGGGGCTTCC	2213
		::::	
Db	871	SerSerSerSerSerAlaProSerSerAlaProSerSerSerThrThrGlu	890
QY	2214	TCTGCTCATGGGCTATACACAGAGCCAGATGTTCAAGGTGGCA-----TCGGGGGTG	2270
Db	891	SerPheSerThrGlyThrThrValThrProSerSerSerValProGlySerThrThr	910
QY	2271	CCCCGGTACACCGTCTGAGTGGCTCTAGCAGACAGAGGTACACTG-----AGCGCTCATAGG	2344
Db	911	GluThrSerSerSerSer-----ThrThrGluThrThrIleValProThrThrThrThr	928
QY	2335	ACGTTCCTGAGAACACACACAGCGCTATGAGGGCGGTTCCGTGCGCTGCACGTGAGA	2384
Db	929	ThrSerValThrThrProSerThrThr-----	937
QY	2385	AGTGGCCAAATGAGCCCAACGCTTGTATATCTCCAGCGGCTTC-----TGACG	2435
Db	938	-----ThrIleThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrThr	955
QY	2436	AAACAGTCACTTTTCCACAAACCTTCCTCGTCTCCCAACATGATCCGACAGGAGAAAC	2495


```

FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571PB9A5663 CRC64;

```

```

Alignment Scores:
Pred. No.: 6.84e-07 Length: 5179
Score: 279.50 Matches: 210
Percent Similarity: 31.85% Conservative: 113
Best Local Similarity: 20.71% Mismatches: 345
Query Match: 5.83% Indels: 347
DB: Gaps: 43

```

US-09-976-674-4 (1-2617) x MUC2_HUMAN (1-5179)

```

QY 18 CCACGGGAGCCCAAGCCGAGCGAGC-----CAG 53
DB 1197 ProProGluAlaSerValProThrClnGluThrCysLysSerCysValCysThrAsnSer 1216
QY 54 CCGCACAGATGACCGCGCCGCTCCAGGTGCAGAGACACTCGTGAGCGGCTCC 113
DB 1217 SerGlnValValCysArgProGluGluLysIleLeuAsnGlnThrGlnAspGlyAla 1236
QY 114 -----GGAGCATCATCCAGCGAGCGGCGAAGTACTCGGCGCTCA 152
DB 1237 PheCysTyrTrpGluIleCysGlyProAsnGluThrValGluLysHisPheAsnIleCys 1256
QY 153 TTGTCAAAAGGCGCCCAAGCACTTCCAGTTGTGCAGAAAGCATAGCTGGGCCCC 212
DB 1257 SerIleThrThrArgProSerThrLeuThrPhe---ThrThrIleThrLeuProThr 1275
QY 213 ACTCCACCGCTCTACTACTGGAATGCATATGCGAGCGAGAGAACTCCCTCTCT 272
DB 1276 ThrProThrSerPheThrThr-----ThrThrThrThrThrThrProThrSer 1291
QY 273 ACTGTGATTCGCCAAGAGGTCCGGAAGAGGCTGCTGCTCTCTGGAAGCAGA 332
DB 1292 SerThrValLeu-----SerThrThrProLysLeuCysCys----- 1303
QY 333 TGTGTGATC-----ATTTCAGGCGCAGCGCCCAACATGGGGTCTACTCTGGGAGAG 386
DB 1304 LeuThrPserAspTrpIleAsnGlnAspHisProSerSerGlySerAspAspGlyAspArg 1323
QY 386 ----- 386
DB 1334 GluProPheAspGlyValCysGlyAlaProGluAspIleGluCysArgSerValLysAsp 1343
QY 387 -----AGCTGCTGAGGAGC----- 401
DB 1344 ProHisLeuSerLeuGluGlnHisGlyGlnLysValGlnCysAspValSerValGlyPhe 1363
QY 402 -----GGAAC----- 407
DB 1364 IleCysLysAsnGlnAspGlnPheGlyAsnGlyProPheGlyLeuCysTyrAspTyrLys 1383
QY 408 -----GCCTGGGGG-----TCCTGCGCATCACTCTCTAGC 437

```

```

DB 1384 IleArgValAsnCysCysTrpProMetAspLysCysIleThrThrProSerProProThr 1403
QY 438 ACTTCACAGCGAGAGTGGCTTCTCCCTTCAGGCGCAGCAACAGCTCTTCACATGTC 497
DB 1404 ThrThr-----ProSerProProProThrThr 1413
QY 498 GCGAGCGGCGGAGAGCGTTTCATGTCGCCCATATAAACCGGTGAATCAAGACC 557
DB 1414 ThrThrLeuProProThrThrThrProSerPro----- 1424
QY 558 AGTGTCAGGCGCCCGATGAGCCCAAAATCGCCCTGCCAGCCCTCTTCTTCTCT 617
DB 1425 -----ProThrThrThrThrThrThrThrProProProThrThrThrProSerPro 1440
QY 618 TCATCATACAGCAGCAGCTGTGGGTGCGCAACATCGAGACAGCGGAGCGCGGCTGA 677
DB 1441 ProIleThrThrThrThrThrProLeuProThrThrThrProSerPro----- 1456
QY 678 CCTTCGCGCACCAAGGTTTATCCAAATGTCGTGATGACCCCAAGTCTCGGGGTGCGCA 737
DB 1457 ProIleSerThrThrThrThrProProProThrThrThrProSer-----Pro 1472
QY 738 CCTTCGTATACAGGAGAGTTCGACCGCTTCACCTGAGTGTGAGTGTGCGCCACAGCT 797
DB 1473 ProThrThrThrThrProSerProProThrThr-----ProSerPro 1486
QY 798 CCTGGGAAGTTACAGAGGCGCTCAAGACGCTGCGAATCCTGTATGAGAGAGTCAAGT 857
DB 1487 Pro-----ThrThrThrThrThr 1493
QY 858 CCGAGTGTGAGTATTCACGTCCTCTCTCGGCTGTGAAGAAAGAGAGAGTCTGT 917
DB 1494 ProProProThrThrThrProSerProProMet-----ThrThrPro 1507
QY 918 ATCGGTACCCAGCAGCAGCAGCAAGAAATCCCAAGATTG-----CCTTGAAC 965
DB 1508 IleThrProProAlaSerThrThrThrLeuProProThrThrThrProSerProProThr 1527
QY 966 TGGCTGACTCCAGACGTACAGCAGCAGCGCAAGATCCTGCACCCAGAGAGAGACTGG 1025
DB 1528 ThrThrThrThrThrProProProThrThrThrThrProSerProProThrThrProIle 1547
QY 1026 TGCAGCCCTTCAGCTCGGTGTCCGAGAGTGAATGACCGAGCGCG----- 1076
DB 1548 ThrProProThrSerThrThrThrLeuProProThrThrThrProSerProProThr 1567
QY 1077 -----GGTGAACCGGAGTGGCAAAATACGCTGGGCCATGTTCTGAGCGGCGCCAGC 1130
DB 1568 ThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrProSer 1587
QY 1131 AGTGGCTCAGCTGCTCTCTCCCGCGCGCTGTTCATCCGAGCAGCAGAGAAATGAG 1190
DB 1588 ProProThrIleThrThrThrThrProProProThrThrThrProSerProProThrThr 1607
QY 1191 AGCAGCGGCTAGCCTTCGAGAGCTGCCAGAGAAATGTCAGCGCATGTGGTACG 1250
DB 1608 ThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrProIleThr 1627
QY 1251 AGGAGTCAACCAAGCTGATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1310
DB 1628 -----ProProThrSerThrThrThrLeuProProThrThrThrProSerProProPro 1645
QY 1311 AGGAGAGAGC-----AGCTCTGCTTCTCCGCGCCCAATGAGAGAGCAG 1358
DB 1646 ThrThrThrThrThrProProProThrThrThrThrProSerProProThrThrThrPro 1665
QY 1359 GCTTCGCAATTGTACAAAGCA-----CCGCGGTTTAAATCCAGGCGTACGATT 1412
DB 1666 SerProProIleThrThrThrThrThrProProProThrThrThrProSerSerProIle 1685
QY 1413 GGAGTGAAGCCCTTCAGCGCGGGAAGATGAATTAAGTCCCATTAAGAGAGAGATTG 1472

```


[illegible]

Search completed: December 12, 2002, 11:49:17
Job time : 72.5 secs

[illegible]

RESULT 2
PCT-US93-07923-2
: Sequence 2, Application PC/TUS9307923
: GENERAL INFORMATION:
: APPLICANT: Morimoto, Chikao
: APPLICANT: Schlossman, Stuart F.
: APPLICANT: Tanaka, Toshaki

```

1 TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
2
3 NUMBER OF SEQUENCES: 16
4
5 CORRESPONDENCE ADDRESS:
6
7 ADDRESSEE: Fish & Richardson
8
9 STREET: 225 Franklin Street
10
11 CITY: Boston
12
13 STATE: Massachusetts
14
15 COUNTRY: U.S.A.
16
17 ZIP: 02110-2804
18
19
20 COMPUTER READABLE FORM:
21
22 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
23
24 COMPUTER: IBM PS/2 Model 502 or 555X
25
26 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
27
28 SOFTWARE: WordPerfect (Version 5.0)
29
30 CURRENT APPLICATION DATA:
31
32 APPLICATION NUMBER: PCT/US93/07923
33
34 FILING DATE: 19930819
35
36 CLASSIFICATION:
37
38 PRIOR APPLICATION DATA:
39
40 APPLICATION NUMBER: 07/934,162
41
42 FILING DATE: 21-AUG-1992
43
44 APPLICATION NUMBER: 07/832,211
45
46 FILING DATE: 06-FEB-1992
47
48 ATTORNEY/AGENT INFORMATION:
49
50 NAME: Fraser, Janis K.
51
52 REGISTRATION NUMBER: 34,819
53
54 REFERENCE/DOCKET NUMBER: 00530/055002
55
56 TELECOMMUNICATION INFORMATION:
57
58 TELEPHONE: (617) 542-5070
59
60 TELEFAX: (617) 542-8906
61
62 TELEX: 200154
63
64 INFORMATION FOR SEQ ID NO: 2:
65
66 SEQUENCE CHARACTERISTICS:
67
68 LENGTH: 759
69
70 TYPE: amino acid
71
72 STRADEDNESS:
73
74 TOPOLOGY: linear
75
76 PCT-US93-07923-2

```

Alignment Scores:		
pred. No.:	1,04e-29	length: 759
Score:	489.50	Matches: 201
Percent Similarity:	36.27%	Mismatches: 126
Best Local Similarity:	22.43%	Mismatches: 330
Query Match:	10.21%	Indels: 241
DB:	5	Gaps: 35

OY	149	CTCAATTGTCACAAAGGGGCCCCAGACTTCCAGTTTGTCAGAAGACGATGTGCGG	208
Db	19	VallLeuIeuAnaLnysGly-----ThraspAspAlaThr	29
OY	209	CCCCACTCCCAAGCCGCCTACTACTCGGAATGCCATATGGCAGCCGAGAACTCCCTC	268
Db	30	AlaAspSerArgLysThrTYrThleu-----	38
OY	269	CTTACTCTGAGATCCCAGAAAGCGTCCGAAAGAGCGCTGCTGCTCTGCTCGGAAG	328
Db	39	-----ThraspyrIleuLysAsnThrTYrArgIeunLysIeuTYrSerLeuArgTrp---	55
OY	329	CAGATGCTGCATCATTTCCACAGGCCAGGCCACCACAGSGGTCTACTCTCGGAGAGAGAC	388
Db	56	-----IleSerAspIseLuIurIeuTYrsgInLysAnsn	68
OY	389	CTGCTG-----AGGAGCGGAAACGCTG	412
Db	69	IleLeuValPheAsnAlaGlutyrGlyAsnSerSerValPheLeucInLysAnserThrPhe	88
OY	413	GGGGCTTTGGGCATCACCTCTACAGATTCCAC---AGCAGAGATGGCCCTTCCTCTTC	469
Db	89	AspAlupIreegLynHisSerIleAsnAspTySerIleSerProAspGlnPheIleLeu	108

OY	470	CAGGCGCAAGCAACGGCTCTTCACCTGTCCGACGGCGGCAAGACGGCTTCATGGTCC	523
Db	109	LeuGIuTyrAsnYrValLysGIuTPrAaHnSserYrThrAlaSerYrAspLeyr	128
OY	530	CCATATGAACCGGTGGAAATCAAGACCCAGTCCAGGCGCCGGATGACCCCAATC	589
Db	129	AspLeuAsnLysArgGlnLeuIleThrGIuGIuAlaIleProAsnAsnThrGIuTPrVal	148
OY	590	TGCCCTCGCAGCCCTGCCCTTC--TTCTCTTCATCAATMAACAGGACCTGTGGGTGCC	646
Db	149	ThrTrpSerProValGIuHisLysLeuAlaTyrValTrpAsnAsnAspLeyrValLys	168
OY	647	AACATCGAGACAGGCGGAGGACGGCGGTGACCTTCGCCACCAAGTTTATCCAAATGC	706
Db	169	IleGIuProAsnLeuProSerYrYrAlaIleThrP-----ThGIuLysGIuAspLle	186
OY	707	CTGGATGACCCCAAGTCTGGCGGTGGGCCACCTCGTCATACAGGAAG--TTGCAC	763
Db	187	IleYrAsn-----GIuIleThrAspTPrValYrGIuGIuGIuValPheSer	202
OY	764	CGCTTCACCTGGTACTGGTGGTCCGCCACAGGCTCTCGGGAAGTTCAGAGGCGCTCAG	823
Db	203	AlaIyrSerAlaLeuTrpTrpSerProAsnGIuThrPhe-----	215
OY	824	ACGCTGGCAATCTGTATGAGAAATGCATAGTCCGAGTGGAGTCAATTCAC-----	877
Db	216	-----LeuAlaTyrAlaGIuPheAsnAspThrGIuValProLeuIleGIuYrSer	232
OY	878	---GTCCCTCTCTCTGGCGGTGAAGAAAGAGAGAGCGACGTCTATCGGTACCCCAAGCA	934
Db	233	PheYrSerAspLysIuSerLeuGIuTyrProLysThrValArgValProLyrProLysAla	252
OY	935	GGCAGCAAGAATCCCAAGATATGCTGTGAACATCGGTGATCCAGATGACAGCCAGGCG	994
Db	253	GIuAlaValAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer	270
OY	995	AAGATCGTCTCGACCCAGGAAGAGACTGGTGCACGCCCTTCAGTCTGCTCCCGAAG	1054
Db	271	serValThrAsnAlaThrSerIleGIuIleThrAlaProAlaSerMetLeuIleGIuLysp	290
OY	1055	GTGGATGACATCGGCAGGCGGGGTGACCCGGGATGGCAATACGGCTGGGCCATGTTC	1114
Db	291	---HisTyrLeuLysAspValThrTrpAlaThrGIuAlaYrIleSer-----	305
OY	1115	CTGGACGGCGCCCGACGAGTGGCTCCAGCTGTCCTCTCCCGCGGCGCTGTATCCCG	1174
Db	306	-----LeuGIuTPrLeu-----	309
OY	1175	AGCACAGAGAAATGAGGAGACGGGATAGGCTTCGCAGAGTGTCCCAAGAAATGCCAG	1234
Db	310	-----ArgArgIleGIu-----	313
OY	1235	CCGATATGGTGG-----TACAGAGAGTCCACACAGTCTGGATCAATGTT	1279
Db	314	AsnYrSerValMetAspLleCysAspYrAspGIuSerSerGIuArgTrp-----	330
OY	1280	CATGACATCTTATCTCTCCCTCCCAATACAGAGGAGAGAGAGAGCTGTCTCTCCG	1339
Db	330	-----	330
OY	1340	GCCAAATGATGCAAGACCGGCTCTTGCCATTTGGTATCAAGTCAAGCGCGTTTAAATCC	1399
Db	331	-----AsnCysIleuValAlaArgGlnHisIleGIuMetSerThr-----	344
OY	1400	CAGGCTACGATTTGGATGAGCCCTTCAGCGCCGGGGAAGATGA-----	1444
Db	345	-----GIuTPrValGIuArgPheArgProSerGIuProHisPheThrLeuAspLey	361
OY	1445	-----TTTAAG	1450
Db	362	AsnSerPheYrLysIleIleSerAsnGIuGIuGIuTyrArgHisIleCysYrPheGIu	381
OY	1451	TGCCCATTAAGGAAGAGATGTCTGTACACAGCGGTAATGGAGAGTTTTGGCGAGGCAC	1510

Db	382	ILASPLYSLYSPCYSTHRRHEIETHRLYSGLYTHRTPLGLVALILEGLYLEGLU	401
OY	1511	GGCTCCAAAGATCTGGGTCATGAGAGACCAAGCTGGGTGTACTTCCAGGCGCAAGAC	1570
Db	402	AlaIeu-----	ThiSerAsp 406
OY	1571	ACGCCCTGGAGCACCACTCTACGTGTACGTATGAGCGCGCGG--	1618
Db	407	-----TYLEIETHTYRILLESASNGILTYRLYSGLYMETPROGLYGLY	421
OY	1619	---GAGTCTGACGGCTCCACCAAGCGCGGCTTCCATAGTGTCTCCATAGCCACAC	1675
Db	422	ArgAsnLeuTYRILYSILEGLINLeuSerAspTYRTHLYSVALTHRCYSLeuSerCysGLU	441
OY	1676	TTTCAGATGTTCCGACACCCACTACACACGTCGTCAGACCGCGCGCTGGGTACAGTCTAC	1735
Db	442	LeuAsnPROGLIARCYSGINLYTYRYSerValSerPheSerLYSGILALILYTYR	461
OY	1736	AAGCTG---ACGCGCCC-----	1750
Db	462	GLINLeuArgCysSerGLYPROGLILEUProLeuTYRTHLeuHisSerSerValAsnAsp	481
OY	1751	-----GACGACGACCCCTGCACACAGCCCGCTTCGGACT	1789
Db	482	LYSGLYLeuArgVALLeuGLuAspAsnSerAlaLeuAspLYS--	495
OY	1790	AGCATGATGAGGAGCGACGACCTGCCCGGATATGTTCTTCCAGATCTGCATATTC	1849
Db	496	---MetLeuGLINAsnVALGln-----MetProSerLYSLYSLeuAspPhe	509
OY	1850	CACACGCGCTCGATGTGTGGGCTC---TACGCGATGATCTACACCCCGACGCTTGACG	1906
Db	510	ILEILEuAsnGLUTHRILYSPhetRPTLYRGLIMETILEuProPheHis---PheAsp	528
OY	1907	CCAGGGAAGAAGCACCCCGCTCTTTGATATGAGAGCGCCCGACGTGACTGTG	1966
Db	529	LYSserLYSTYRProLeuLeuAspVALTYRALGLYProCysSerGLINLYSAla	548
OY	1967	AATTAACCTCTTCAAGGACATCAAGTACTTGGCGCTCAAC-----ACATGGCGCTCC	2017
Db	549	AspHRVALPhe-----ArgLeuAsnTTPAlaTHRTYRLeuAlaSer	562
OY	2018	CTGGGCTACGCGCGTGTGTG---ATTGACGGCAGGGGCTCTCTCAGCAGGCGCTTGG	2074
Db	563	ThrgLInsnIleIleVALIaSerPheAspGLYARGGLYserGLYTyrGLInGLYAspLYS	582
OY	2075	TTTCCAGAGGCGCTCAAAAACCAATATGGCCAGCATGAGATGAGACCAAGGTGAGGC	2134
Db	583	ILEMetHisAlaIleAsnAspARGLeuGLYTHRPhEGLIValGLINAspGLINleGLIVala	602
OY	2135	CTG---CAGTTCGCGGGCGGAGATATGGCTCATGACCTGAGCGAGTGGCCATCAT	2191
Db	603	AlaArgGLINPhe-----SerLYSMetGLYPheVALAspAsnLYSArgILEAlaIleTPR	620
OY	2192	GGCTGTCTCTACGGGGGCTCTCTGTGCATGAGGGCTATCCAAAGCGCGAGTGTTC	2251
Db	621	GLYTPRserTYRGLYGLYTYRValTHNserMetValLeuGLYserGLYserGLYAlaPhe	640
OY	2252	AAGTGGCCATCGGGGTGCCCGCGTACCCTGTGATGGCTTACGACACAGGTACT	2311
Db	641	LYSCysGLYIleAlaVALAlaProVALserARGTYRGLIUTYRAspSerValTYRTHR	660
OY	2312	GAGGCTATCATGAGACGTC-----CTGAGAACACACACAGCGCTATGAGCGGGCTCC	2365
Db	661	GLUTArgTYRMetGLYLeuProTHRProGLUAspAsnLeuAspHisTYRArgAsnSerTHR	680
OY	2366	GTGGCGCTGACGTGAGAAAGTGCACATGAGACCCCAACCGTGTCTATCTCCACGAC	2425
Db	681	VALMetSerArgAlaGLUAsnPhe-----LYSGINValGLIUTYRLeuLeuILEHisLY	698
OY	2426	TTTCTGAGCAAAAAGTGCATTTTTCACACAAATCTCTCGTGTCCCAACTGACCA	2485

```
D6      699   ThrAlaAspaSPasValHisPheGlnclnseralacInlleserLySAlalaauVaLasp    718
QY      2466 GCAGGGAAACTTACCAcAcCTCCATCATTCACCccAACAgAACAcaTAATTGCctGCCccc    254
        ||| :||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D6      719 ValGIyValaSPpheglnAlameTrpyrThrSpIunspHslgllelaaserSer    738
QY      2546 GAGTGCGGCGAGCAGCTATGAAGTAGCACGTgcTGCAtTTCTTACAGGAA    2593
        ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D6      739 ThrlAhIsGlNhlSlletYTThrhISmetberIslshellelysgln    754

RESULT 3
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANTS: Garlin-Chese, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELFE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: IUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEX: (212) 838-3884
INFORMATION FOR SEQ ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-3

Alignment Scores:
Pred. No.:          Length:       766
Score:              Matches:         201
                    Percent Similarity: 26.27%
                    Mismatches:        330
                    Query Match:     10-.21% Indels:      241
DE:                  Gaps:            35

US-09-976-674-4 (1-2617) x US-08-230-491A-3 (1-766)
QY      149 CTCATNTGTCACCAAGCGCCCCCAGAGACTTCAGTTGTtGCaAGAAGAGcatGaGtcTcgGg    208
        ::::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D6      26 ValLeuleduanlsylsGLY-----ThraSPasPaLaThr    36
QY      209 CCCACCTCCcACCGCcCTTACTACTCTGGGAATGCCATATGGcAGcCCAGAGAACTCCCTC    268
        ||| ::: ||| || | | | | | | | | | | | | | | | | | | | | | | | |
D6      37 AlaasPeSaTaaglysThryTHrleuu-----'                45
QY      269 CTCTACTGTgaGaatTCCCAGAGAGGTCCGGAAGAGAGGCTGTGCTGCTCTGCTCTGtgGAG    328
        ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D6      46 -----ThrasPyTLeuLYsaSnThTyrtArgLeuLYseurTYrsreLtauArGrtp---    62
```

OY	329	CAGATGCGGATCATTTCTTCCAGGCGACCGCCACCATGGGGTACTGCTGGGAGAGG	388
Db	63	-----11esAspHisGluTyrLeuTyrGlnGluAsnAsn	75
OY	389	CTGCTG-----AGGAGCGGAAGCGCTG	412
Db	76	1leuValPheAsnAlaGluTyrGluYAsnSerSerValPheGluGluAsnSerThrPhe	95
OY	413	GGGGCTTCGGCATACCTCCCTACAGCTTCAC--AGCAGAGTGGCTCTTCCTTC	469
Db	96	AspGluPheGlyHisSerIleAsnAspTyrSerIleSerProAspGluPheIleLeu	115
OY	470	CAGCGCCACCAACACCTTCCTCCACTGCGCGGAGCGGCGCAAGAACGGCTCAATGGTGC	529
Db	116	LeuGluTyrAsnTyrValLysGlnTPrpGlnHisSerTyrThrAlaSerTyrAspIleTyr	135
OY	530	CGTAAACCGCTGGGAATCAAGACCCATGCTGACGGCCCGGATGAGCCCAAAATC	589
Db	136	AspLeuAsnLysArgGlnLeuIleThrGluGluArgIleProAsnAspThrGlnTPrpAla	155
OY	580	TGCCCTGCCGACCTGCTTC--TTCTCTTCATCAATPACAGCAGCCTGTGGTGGCC	646
Db	156	ThrTPrpSerProValGlnHisLysLeuAlaTyrValTPrpAsnAsnAspIleTyrValLys	175
OY	647	AACATCGAGACAGGCGAGGAGGCGGGCGCTGACCTTGGCCACCAAGCTTATCCAAATGC	706
Db	176	IleGluProAsnLeuProSerTyrArgIleThrTPrp----ThrLysGluAspIle	193
OY	707	CTGGATGACCCCAAGTCTGGGGGTGGTGCCACCTTGTCATACAGAGAAAG--TTTCAC	763
Db	194	IleTyrAsn-----GlyIleThrAspTPrpValTyrGluGluGluValPheSer	209
OY	764	CGCTTACCTGGGTACTGTGGTGCGCCACAGCCTCTGGGAAGTTACAGAGGCGCTCAAG	823
Db	210	AlaTyrSerAlaLeuTPrpTPrpSerProAsnGlyThrPhe-----	222
OY	824	ACGCTGGCAATTCGTATGAGAAAGTCAGATGATCCGAGGTGGAGGTCAATTAC-----	877
Db	223	-----LeuAlaTyrAlaGlnPheAsnAspThrGluValProLeuIleGluTyrSer	239
OY	878	--GTCCTCTCTCTGGCTGAGAAAGAAAGAAACAGCAGCTCGATGCGTACCCAGACA	934
Db	240	PheTyrSerAspGluSerLeuGlnTyrProLysThrValArgValProTyrProLysAla	259
OY	935	GGCAGACAGAATCCCAAGATGCTCTTGAAATCTGGCTGAGTCCAGACTGACAGCGAGGC	994
Db	260	GlyAlaValAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer	277
OY	995	AAGATCGCTGTGACCCAGAGAAAGAGAGCTGTGCACCTTCAGCTCGCTGTCCCGAAG	1054
Db	278	SerValThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp	297
OY	1055	GTGGATGATCAATCGCCAGAGGCGGGGTGGACCCGGGATGGCAAAATACGCCGTGGCCATGTTC	1114
Db	298	---HisTyrLeuCysAspValThrTPrpAlaThrGlnGluArgIleSer-----	312
OY	1115	CTGAGCGGCGCCAGCAGTGGCTTCAGACTGTCTCTCCCGCCGCTGTTCATCCG	1174
Db	313	-----LeuGlnTPrpLeu-----	316
OY	1175	AGCACAGAGATGAGGAGACGGGCTAGCTTGCAGAGCTGTCCCAAGAAATGCAG	1234
Db	317	-----ArgArgIleGln-----	320
OY	1235	CCGATGTGGTG--TACAGAGAGGTCAACCAAGCTGTGATCAATGTT	1279
Db	321	AsnTyrSerValMetAspIleCysAspTyrAspGluSerSerGlyArgTPrp-----	337
OY	1280	CATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTGTGCTTCTCCGC	1338
Db	337	-----	337
OY	1340	GCCAAATGAATCAAGACCGGCTTCGCAATTGTTCAAAGTCAACGGCGTTTAAATATCC	1399

```

Db 338 -----AsnCySLeuValAlaArgGlnHisIleGluMetSerThrThr----- 351
QY 1400 CAGGGCTACGATTGGATGAGCCCTTCAGCCCGGGGAGATGAA----- 1444
Db 352 -----GlyTrpValGlyArgPheArgProSerGluProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAG 1450
Db 369 AsnSerPheThrIleIleSerAsnGluGluGlyTrpArgHisIleCysTyrPheGln 388
QY 1451 TGCCCATTAAGGAAGATTGCTCTGACAGCGGTGAATGGAGGATTGGCGAGCAC 1510
Db 389 IleAspLysAspCysThrPheIleThrLysGlyThrTrpGluValIleGlyIleGlu 408
QY 1511 GGCCTCAAGATCTGGGTCATGAGAGACCAAGCTGGTACTTCCAGGGCAGCAGAGAC 1570
Db 409 AlaLeu-----ThrSerAsp 413
QY 1571 ACGCCGCTGGAGCACACCTCTACGTGTCAGTATGAGCGCGCGG----- 1618
Db 414 -----TyrLeuTyrTrpIleSerAsnGluTyrLysGlyMetProGlyGly 428
QY 1619 ---GAGATCGTACGCTCACCACGCGCGGCTTCTCCATAGCTGCTCATGAGCCAGAAC 1675
Db 429 ArgAsnLeuTyrLysIleGlnLeuSerAspTyrThrLysValThrCysLeuSerCysGlu 448
QY 1676 TTCGACATGTTCTGACGACCTACAGACGCGTGAGACGCGCGCGCTGACAGCTAC 1735
Db 449 LeuAsnProGluArgCysGlnTyrTyrSerValSerPheSerLysGluAlaLysTyrTyr 468
QY 1736 AAGCTG-----ACGGGCCC----- 1750
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyrThrIleuHisSerSerValAsnAsp 488
QY 1751 -----GACGACGACCCCTGCACAGACGCGCGCTTCTGGGCT 1789
Db 489 LysGlyLeuArgValLeuGlnAspAsnSerAlaLeuAspLys----- 502
QY 1790 AGCATGATGAGGAGCAGCAGCTGCCCCCGGATATGTTCTCCAGAGATTTCCATTTC 1849
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGGGCTCGATGTCGGGCTC---TAGCGATGATCTCAAGCGCGCGCTTGAG 1906
Db 517 IleIleLeuAsnGlnTyrLysPheTrpTyrGlnMetIleLeuProHis---PheAsp 535
QY 1907 CCAGGAGAAGAGACACCCACCGCTCTTTGTATATGAGAGCGCCAGAGTGCAGCTGCTG 1966
Db 536 LysSerLysLysTyrProLeuLeuAspValTyrIleGlyProCysSerGlnLysAla 555
QY 1967 AATACCTCTTCAAAAGCATCAAGTACTTGGCGCTCAAC-----ACACTGGCCTCC 2017
Db 556 AspThrValPhe-----ArgLeuAsnTrpAlaThrTyrLeuAlaSer 569
QY 2018 CTGGGCTACGCGCGGTTGTG---ATTGAGGAGGAGGCGCTCTGTCAGCGGCGTTGG 2074
Db 570 ThrGlnAsnIleIleValAlaSerPheAspGlyArgLysGlyTyrGlnGlyAspLys 589
QY 2075 TTCGAAGGGGCGCTGAAAAACCAATGAGCGAGTGCAGATCGAGAGCAGGAGTGGAGGC 2134
Db 590 IleMetHisAlaIleAsnArgArgLysGlyThrPheGluValGluAspGlnIleGluAla 609
QY 2135 CTG---CAGTTCGTGGCGGAGAGATGAGCTTCATGACCTGAGCCGAGTTGCCATTCAT 2191
Db 610 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTrp 627
QY 2192 GGCAGTCTACGGGGGCTCTCTCGCTCATGGGGCTAAATCCACAAAGCCCGAGGTGTC 2251
Db 628 GlyTrpSerTyrGlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValPhe 647
QY 2252 AAGGTGGCATGCGGGGTGCCCGGTCAACGCTTGATGGCTTACGACAGAGGTACACT 2311

```

```

Db 648 LysCysGlyIleAlaValAlaProValSerArgTrpGluTyrTyrAspSerValTyrThr 667
QY 2312 GAGCGCTACATGAGACGTC-----CTGAGAACACACGACGCGCTATGAGCGGTTCC 2365
Db 668 GlnArgTyrMetGlyLeuProThrProGlnAspAsnLeuAspHisTyrArgAsnSerThr 687
QY 2366 GTGGCCCTGACGTCGAGAAAGCTGCCCCAATGAGGCCCAACCGCTGTATCTTCACAGGC 2425
Db 688 ValMetSerArgAlaGluAsnPhe-----LysGlnValGluTyrLeuLeuIleHisGly 705
QY 2426 TTCCTGACGAAAAACGTGCACCTTTTTCACACAAACTTCTCGTCCCAATGATCCGA 2485
Db 706 ThrAlaAspAspAsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAsp 725
QY 2486 GCAGGAACCTTACCAAGCTCCACATCTACCCCAACGAGACACACAGTATTCGCTGCC 2545
Db 726 ValGlyValAspPheGlnAlaMetTrpTyrThrAspGluAspHisGlyIleAlaSerSer 745
QY 2546 GAGTCGGGCGAGCACATGAACTGACGTTGCTGCTGCTTCTACAGAA 2593
Db 746 ThrAlaHisGlnHisIleTyrThrHisMetSerHisPheIleLysGln 761

```

RESULT 4

US-08-619-280A-3

Sequence 3, Application US/08619280A

Patent No. 5767242

GENERAL INFORMATION:

APPLICANT: Zimmermann, Rainer; Park, John E.;

APPLICANT: Retlig, Wolfgang; Old, Lloyd J.

TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felle & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/619,280A

FILING DATE: 18-MARCH-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/230,491

FILING DATE: 20-APRIL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5767242man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5330.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 638-3884

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 766 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-619-280A-3

Alignment Scores:

Pred. No.:	1,04e-29	Length:	766
Score:	489.50	Matches:	201
Percent Similarity:	36.27%	Conservative:	124
Best Local Similarity:	22.43%	Mismatches:	330
Query Match:	10.21%	Indels:	241
DB:	1	Gaps:	35

US-09-976-674-4 (1-2617) x US-08-619-280A-3 (1-766)

QY 149 CTCATGTTCAACAGGCGCCACAGACTTCAGATTGTGACAGAGAGCATGAGTGG 208
: : : : :
Db 26 ValLeuLeuasnLysGly-----ThrAspAspAlaThr 36
QY 209 CCCCACATCCACCGCCTTACTACTGGAATGCCAATATGACCGACAGACACTCCCTC 268
: : : : :
Db 37 AlaAspSerAlaGlySthIrrTyThrLeu----- 45
QY 269 CTCACACTGATGCCAAGAAAGTCCGGAAGAGGCTGCTGCTCTCTCTGAGAG 328
: : : : :
Db 46 -----ThrspryIleuLysAsnThrrTyArGleuLysLeuTySerLeuAlaGrrp--- 62
QY 329 CAGATGCTGATCATTTCCAGGCGCACCACCATGGGCTACTCTCGGAGAGAGAG 388
: : : : :
Db 63 -----IleSerAspHisGluTyrrLeuTyrrLysGlnGluAsnAsn 75
QY 389 CTGCTG-----AGGAGCGGAAAGCGCTG 412
: : : : :
Db 76 IleLeuValPheAsnAlaGluTyrrGlyAsnSerSerValPheLeuGluAsnSerThrPhe 95
QY 413 GGGGCTTCGGCATCAGCTCTACGACTTCAC---AGCGAGATGGCTCTTCTCTTC 469
: : : : :
Db 96 AspGluPheGlyHisSerIleAsnAspTyrrSerIleSerProAspGlyGlnPheIleu 115
QY 470 CAGCGCAGCAACAGCCTTTCACACTGCGGAGCGGCGCAAGAACGGCTTCATGCTGCC 529
: : : : :
Db 116 LeuGluTyrrAsnTyrrLysGlnTrpAlaHisSerTyrrThrAlaSerTyrrAspIleTyrr 135
QY 530 CCTATGAACCCGCTGGAAATCAGACCCAGTCTCAGGCGCCGCGAGTGCACCAAAATC 589
: : : : :
Db 136 AspLeuasnLysArgGlnLeuIleThrGluGluArgIleProAsnAsnThrGlnTrpVal 155
QY 590 TGGCCCTCGACACCTGCTTC---TTCCTCTATCATATAACAGACGCTGCTGGTGGCC 646
: : : : :
Db 156 ThrTrpSerProValGlyHisLysLeuAlaTyrrValTrpAsnAspIleTyrrLys 175
QY 647 AACATCGACAGAGCGGAGGCGCGCTGACCTTCGCCACCAAGGTTATCCAAATGTC 706
: : : : :
Db 176 IleGluProAsnLeuProSerTyrrArgIleThrTrp-----ThrGlyLysGluAspIle 193
QY 707 CTGATGACCCCAAGTGTGGGGGTGGCCACCTTCATACAGAGAG---TTGAC 763
: : : : :
Db 194 IleTyrrAsn-----GlyIleThrAspTrpValTyrrGluGluValAlaPheSer 209
QY 764 CGCTTCACTGGTACTGGTGGTGGCCACAGCCTCGGAGGTTGAGAGGCTCAG 823
: : : : :
Db 210 AlaTyrrSerAlaLeuTrpTrpSerProAsnGlyThrPhe----- 222
QY 824 AGCCTGCGAAATCTGTATGAGAAAGTCGATGATCCGAGGTGAGATCATTCAC----- 877
: : : : :
Db 223 -----LeuAlaTyrrAlaGlnPheAsnAspThrGluValProLeuIleGluTyrrSer 239
QY 878 ---GTCCCTCTCTCGGCTAGAAAGAAAGAGACGACCTGATCGGTACCCAGACA 934
: : : : :
Db 240 PheTyrrSerAspGluSerLeuGlnTyrrProLysThrValAlaArgValProTyrrProLysAla 259
QY 935 GGCAGCAAGATCCCAAGATTTGCTTGAACATGCTGATGTTCCAGTCCAGGAGGCG 994
: : : : :
Db 260 GlyAlaValAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer 277
QY 995 AAGATGCTCGACCCAGAGAGAGAGAGCTGGTCAAGCCCTTCAGCTCGCTGCCAGAG 1054
: : : : :
Db 278 SerValThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp 297
QY 1055 GTGGAGTACATCCGACGCGCGGTGAGCCGGATGAGCAATACGCTTGGCCATCTTC 1114
: : : : :
Db 298 ---HisTyrrLeuLysAspValThrTrpAlaThrGlnGluArgIleSer----- 312
QY 1115 CTGAGACGCGCCAGAGAGGCTGCTCAGCTGCTCTCTCCCGCGCGCTTCATCCCG 1174
: : : : :
Db 313 -----LeuGlnTrpLeu----- 316

QY 1175 AGCAGAGAAATGAGAGACAGCGGCTAGCCTCTCCAGAGCTGTCCCAAGATGTCAG 1234
: : : : :
Db 317 ----- 320
QY 1235 CCGTATGTGGTG-----TACGAGAGGTCAACAGTGTGATCAATGTT 1279
: : : : :
Db 321 AsnTyrrSerValMetAspIleCysAspTrpAspLysSerGlyArgTrp----- 337
QY 1280 CATACATCTTTCATCCCTTCCCAATCCAGAGGAGAGAGAGCTGCTTCTCCGC 1339
: : : : :
Db 337 ----- 337
QY 1340 GCCAATGATGACAGACCGGCTTCTGCCATTTGACAAGTACCGCGTTTAAATCC 1399
: : : : :
Db 338 -----AsnGlyLeuValAlaArgGlnHisIleGluMetSerThrTrp----- 351
QY 1400 CAGGCTACGATTTGAGTGAAGCCCTTCAGCCCGGAGAGATGAA----- 1444
: : : : :
Db 352 -----GlyTrpValGlyArgPheArgProSerGluProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAAG 1450
: : : : :
Db 369 AsnSerPheTyrrLysIleIleSerAsnGluGluGlyTyrrArgHisIleCysTyrrPheGln 388
QY 1451 TGGCCCATTAAGAGAGATTTGCTGACACGCGGTGAATGGAGTTCGCGAGGCGAC 1510
: : : : :
Db 389 IleAspLysLysAspCysThrPheIleThrLysGlyThrTrpGluValIleGlyIleGlu 408
QY 1511 GCCTCCAGATCTGGGTCAATGAGAGACCAAGCTGGTACTTCAGGCGACCAAGGAC 1570
: : : : :
Db 409 AlaLeu-----ThrSerAsp 413
QY 1571 ACGCGCTGAGACACACCTCTACGTCAGTACATGAGGCGCGCGC----- 1618
: : : : :
Db 414 -----TyrrLeuTyrrIleSerAsnLysGlnGlyMetProGlyGly 428
QY 1619 ---GAGATCGTAGCCCTCACACGCGCGGCTTCCCATGACTGCTTCATGACAGACAGAC 1675
: : : : :
Db 429 ArgAsnLeuTyrrLysIleGlnLeuSerAspTrpThrLysValThrCysLeuSerCysGlu 448
QY 1676 TTGACATGTTGTCAGACCCCTACAGAGCGTGAACAGCGCGCGCTGCGACAGTCTAC 1735
: : : : :
Db 449 LeuAsnProGluArgCysGlnTyrrTyrrSerValSerPheSerLysGluAlaLysTyrr 468
QY 1736 AAGCTG-----AGCGGCGCC----- 1750
: : : : :
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyrrThrLeuHisSerSerValAsnAsp 488
QY 1751 -----GACGAGACCCCTGACAGAGCGCGCTTCTGGGCT 1789
: : : : :
Db 489 LysGlyLeuArgValLeuGluAspAsnSerAlaLeuAspLys----- 502
QY 1790 AGCATGATGAGAGCGACGACACTGCCCCCGGATTAATGTTCTCTCAGAGATCTTCATTTTC 1849
: : : : :
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGCGCTCGGATGTCGCGCTC---TACGCGATGATTAACAAGCCCGCGCTTCAG 1906
: : : : :
Db 517 IleIleLeuAsnGluThrLysPheTrpTrpGlnMetIleLeuProProHis---PheAsp 535
QY 1907 CCAAGGAAGAGACCCACGCTCTTGTATATGAGAGCCCGCCAGGTGACGTGGTG 1966
: : : : :
Db 536 LysSerLysLysTyrrProLeuLeuLeuAspValTyrrAlaGlyProCysSerGlnLysAla 555
QY 1967 AATAACTCTTCAAAAGCATCAAGTACTTCGCGCTCAAC-----ACACTGCGCTTC 2017
: : : : :
Db 556 AspTrpValPhe-----ArgLeuAsnTrpAlaThrTyrrLeuAlaSer 569
QY 2018 CTGAGCTACGCGTGGTGG---ATTGACGAGAGGCGCTCTGTCAGCAGGCGCTTCGG 2074
: : : : :
Db 570 ThrGluAsnIleIleValAlaSerPheAspGlyArgGlySerGlyTyrrGlnGlyAspLys 589


```

0Y 2075 TTCACAGGGGCGCCGAGAAACCAATGAGGGGACAGGTGATGAGACACAGCGTGGAGGC 2134
Db 590 ILeMethIsaIaIleasnaIghaIleuGIyIthrPheGluValaIGluAspInIleIduLa 609
0Y 2135 CTG---CAGTTCGTGGCCGAGAAAGTAAATGAGCTTCATGACCTAGCCGAGGTGCCATTCAT 2191
Db 610 AlaArgInPhe-----SerLysMetGIyPheValAspAsnIyArgIleAlaIleTrp 627
0Y 2192 GCGTGTCTTCACGGGGGCTTCCTCTGCTGCATCAGTGGGCTATATCCACAAGCCCAAGGTGTC 2251
Db 628 GLytrPserIyGIyGIyValIthrSerMetValIleuGIySerGIyAlaPhe 647
0Y 2252 AAGGTGGCCATCCGGGGGTGCCCGGGTCCACCGTCCATGATGCGCTACGACACAGGTAACCT 2311
Db 648 LysIyScIyIleAlaValaIaIProValIleSerAlaGlytrGIyIyIyAspSerValIyIthr 667
0Y 2312 GAGCGCTACATGAGACGTC-----CCTGAGAACACACAGCAGCGGTATGAGGCGGGGTCC 2365
Db 668 GLuArgIyIyMetGIyLeuProIthrProGIuIyAspAsnIleuAspHisIyIyArgAsnSerIthr 687
0Y 2366 GTGGCCCTGCACGTGAGAAAGTCGCCAATGAGCCCAACCGCTTGCTATATCCTCCAGCGC 2425
Db 688 ValMetSerArgIaGIuAsnPhe-----LysGIuIyAlGIyIyLeuIleuIleHisIy 705
0Y 2426 TTCTGACACAAAACGTCGCACTTTTCACACAAACTTCTGTGCTCCCACTATCCGA 2485
Db 706 ThrAlaAspAsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaIleuValaIsp 725
0Y 2486 GCAGGAGAACCTTACACGACGTCGCAATCTACACCCACAGAGACACAGATTCGCTGCCGCC 2545
Db 726 ValGIyAlaIspPheGlnAlaMetIyIyIyIyAspGIuIyAspHisGIyIyAlaIaser 745
0Y 2546 GAGTCGGGCGAGCAGCATATGAAGTCACGTTGCTGCACTTTCTACAGAA 2593
Db 746 ThrIaHisGlnHisIyIyIyIyIyIyIyIyIyIyIyIyIyIyGln 761

RESULT 5
US-08-940-391-3
: Sequence 3, Application US/08940391
: Patent No. 5965373
: GENERAL INFORMATION:
: APPLICANT: Zimmermann, Rainer; Park, John E.;
: TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felle & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
: COUNTRY: USA
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/940,391
: FILING DATE: 01-OCT-1997
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/619,280
: FILING DATE: 18-MARCH-1996
: APPLICATION NUMBER: 08/230,491
: FILING DATE: 20-APRIL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 5965373man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: IUD 5330.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 688-9200

```

TELEFAX: (212) 838-3884	TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:	INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:	SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids	LENGTH: 766 amino acids
TYPE: amino acid	TYPE: amino acid
TOPOLOGY: Linear	TOPOLOGY: Linear
US-08-940-391-3	US-08-940-391-3
Alignment Scores:	Alignment Scores:
Pred. No.: 1,04e-29	Pred. No.: 1,04e-29
Score: 489.50	Score: 489.50
Percent Similarity: 36.27%	Percent Similarity: 36.27%
Best Local Similarity: 22.43%	Best Local Similarity: 22.43%
Query Match: 10.21%	Query Match: 10.21%
DB: 2	DB: 2
US-09-976-674-4 (1-2617) x US-08-940-391-3 (1-766)	US-09-976-674-4 (1-2617) x US-08-940-391-3 (1-766)
149 CTCATGTCACACAAAGGCGCCAGACTTCAGTTTGTGACAGAGCGATGAGTCTGGG	149 CTCATGTCACACAAAGGCGCCAGACTTCAGTTTGTGACAGAGCGATGAGTCTGGG
26 ValIleuLeuAsnLysGly	26 ValIleuLeuAsnLysGly
209 CCCACCCACCGGCTCTACTACTGGGATGGCATATGACACCCAGAGAACTCCCTC	209 CCCACCCACCGGCTCTACTACTGGGATGGCATATGACACCCAGAGAACTCCCTC
37 AlaaspSerArgLysThrTyrThrLeu	37 AlaaspSerArgLysThrTyrThrLeu
269 CTCCTACTGTGAGATCCCAAGAAAGGTCCGGAAGAGGCTGCTGCTCTGTCTGGAG	269 CTCCTACTGTGAGATCCCAAGAAAGGTCCGGAAGAGGCTGCTGCTCTGTCTGGAG
46 -----ThrAspTyrLeuLysAsnThrTyrArgLeuLysLeuTyrSerLeuArgTyr	46 -----ThrAspTyrLeuLysAsnThrTyrArgLeuLysLeuTyrSerLeuArgTyr
329 CAGATGCTGGATCATTTCCAGGCGACGCCACCACATGGGGTCTACTCTCGGAGAGAG	329 CAGATGCTGGATCATTTCCAGGCGACGCCACCACATGGGGTCTACTCTCGGAGAGAG
63 -----IleSerAspHisGluTyrLeuTyrLysGlnGluAsnAsn	63 -----IleSerAspHisGluTyrLeuTyrLysGlnGluAsnAsn
389 CAGCGT-----AGGACGGCGAAAGCGCTG	389 CAGCGT-----AGGACGGCGAAAGCGCTG
76 IleIleuValIlePheAsnAlaGluTyrGlyLysAsnSerSerValPheLeuGluAsnSerThrPhe	76 IleIleuValIlePheAsnAlaGluTyrGlyLysAsnSerSerValPheLeuGluAsnSerThrPhe
413 GGGGCTTCGGCAGACACCTCCAGACTTCACAC-----AGCGAGAGTGGCGCTTCCTTC	413 GGGGCTTCGGCAGACACCTCCAGACTTCACAC-----AGCGAGAGTGGCGCTTCCTTC
96 AspGluPheGlnLysSerIleAsnAspArgTyrSerIleSerProAspGlnGlnPheIleLeu	96 AspGluPheGlnLysSerIleAsnAspArgTyrSerIleSerProAspGlnGlnPheIleLeu
470 CAGGCCAGCAACAGCGCTTCCTCCACTGTGCGAGCGCGGCAAGAAAGCGCTTCAGTGTC	470 CAGGCCAGCAACAGCGCTTCCTCCACTGTGCGAGCGCGGCAAGAAAGCGCTTCAGTGTC
116 LeuGluTyrAsnTyrValLysGlnThrPyrGlnHisSerTyrThrAlaSerTyrAspIleTyr	116 LeuGluTyrAsnTyrValLysGlnThrPyrGlnHisSerTyrThrAlaSerTyrAspIleTyr
530 CCTATGAACCCGCTGGAATACAACCCAGAGCTCAGGCGCCGGATGACCCCAAAATC	530 CCTATGAACCCGCTGGAATACAACCCAGAGCTCAGGCGCCGGATGACCCCAAAATC
136 AspleuAsnLysArgGlnIleuIleThrGlnGluArgIleProAsnAsnThrGlnTyrVal	136 AspleuAsnLysArgGlnIleuIleThrGlnGluArgIleProAsnAsnThrGlnTyrVal
590 TGGCCCTGGCGACCGCTTCCTC-----TTTCCTTCATCATTAACAGCGACGCTGGGTGGC	590 TGGCCCTGGCGACCGCTTCCTC-----TTTCCTTCATCATTAACAGCGACGCTGGGTGGC
156 ThrTyrSerProValGlyHisLysLeuAlaTyrValIleThrPheAsnAspIleTyrValLys	156 ThrTyrSerProValGlyHisLysLeuAlaTyrValIleThrPheAsnAspIleTyrValLys
647 AACATCGAGACAGCGGAGAGCGCGCGCTGCACTTCGCCACCAAGAGTTATCCAAATGC	647 AACATCGAGACAGCGGAGAGCGCGCGCTGCACTTCGCCACCAAGAGTTATCCAAATGC
176 IleGluProAsnLeuProSerTyrArgIleThrTyr-----ThrGlyLysGluAspIle	176 IleGluProAsnLeuProSerTyrArgIleThrTyr-----ThrGlyLysGluAspIle
707 CTGGATGACCCCAAGTCTGGCGGTGTGGCCACTTCCTGTCATACAGGAAG-----TTGAC	707 CTGGATGACCCCAAGTCTGGCGGTGTGGCCACTTCCTGTCATACAGGAAG-----TTGAC
194 IleTyrAsn-----GlyIleThrAspThrValTyrGlnGlnGluValPheSer	194 IleTyrAsn-----GlyIleThrAspThrValTyrGlnGlnGluValPheSer
764 CGCTTCACCTGGGTACTGTGGGTGGCGCCACAGCTCCGCGGAAGGTCCAGAGGGCTCAG	764 CGCTTCACCTGGGTACTGTGGGTGGCGCCACAGCTCCGCGGAAGGTCCAGAGGGCTCAG
210 AlaTyrSerAlaLeuThrTyrTyrSerProAsnGlyThrPhe-----	210 AlaTyrSerAlaLeuThrTyrTyrSerProAsnGlyThrPhe-----
824 ACGCTCGAATCTGTATAGAGGAAGTCGATGCTGAGTGGAGGTCAATTCAC-----	824 ACGCTCGAATCTGTATAGAGGAAGTCGATGCTGAGTGGAGGTCAATTCAC-----
223 -----LeuAlaTyrAlaGlnPheAsnAspPheGlnValProLeuIleGluTyrSer	223 -----LeuAlaTyrAlaGlnPheAsnAspPheGlnValProLeuIleGluTyrSer
878 ---GTCCCTCTCTCGGCGCTGAGAAAGAAAGAGAGCGAGCTGTATCGGTACCCAGAGCA	878 ---GTCCCTCTCTCGGCGCTGAGAAAGAAAGAGAGCGAGCTGTATCGGTACCCAGAGCA
240 PheTyrSerAspGlnSerIleGlnTyrProLysThrValArgValProTyrProLysAla	240 PheTyrSerAspGlnSerIleGlnTyrProLysThrValArgValProTyrProLysAla

OY	935	GGGAGCAAGATCCCAAGATTGGCTTGAACACTGGCTGAGTTCCAGACTGACGCCAGCC	994
Db	260	GLYAlaValAsnProThrVal-----LysPhePheValAlaAsnHisPheLeuSer	277
OY	995	AAGATCTCTTCAGCCACGGAGAAAGACAGCTGGTGCACCCCTTAGCTCGCTGTTCCCGAAG	1054
		:::	
Db	278	SerValThrAsnAlaThrSerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp	297
		:::	
OY	1055	GTGAGATACATCGCCACAGGCCGGGTGGACCCGGAGATGGCAATACCGCTGGGCCATGTC	1114
Db	298	---HisTyrLeuCysAspValThrTrpAlaThrGlnIleArgIleSer-----	312
OY	1115	CTGGACCGGGCCCGACGACAGTGCACAGCTGTCCTGTCGCCCGGCTGTTCATCCG	1174
Db	313	-----LeuGlnTrpLeu-----	316
OY	1175	AGCACACAGAAATGAGACAGCCGGTACCTCTGCCAGAGCTGTGCCACGAAATGCTCAG	1234
Db	317	-----ArgArgIleGln	320
OY	1235	CCGATATGGTG-----TACAGAGAGGTACCCACAGTGTGGATCAATGTT	1279
Db	321	AsnTyrSerValMetAspIleCysAspTyrAspGlnSerSerGlyArgTrp-----	337
OY	1280	CATGACATCTTATCCCTTCCCAATGACAGAGAGACAGACTGTGCTTTCGCG	1339
Db	337	-----	337
OY	1340	GCCAAATATGCAAGACCGGCTCTGCCATTTGTACAAAGTACCGCGCTTTAAATCC	1399
Db	338	-----AsnCysLeuValAlaArgGlnHisIleGluMetSerThrThr-----	351
OY	1400	CAGGGCTACAGATTGGAGTGTGACCCCTTCAGCGCCGGGAGATGAA-----	1444
Db	352	-----GlyTrpValGlyArgPheAspArgProSerGlnProHisPheThrLeuAspIle	368
OY	1445	-----TTTAG	1450
Db	369	AsnSerPheTyrIleIleIleSerAsnGlnGluGlyTyrArgHisIleCysTyrPheGln	388
OY	1451	TGCCCACTTAAGAAAGAGATTGCTCTGCACAGCGGTGAAGAGAGTTTGGCGAGAC	1510
Db	389	IleAspIleCysAspCysThrPheIleThrLysGlyThrTrpIleValIleGlyIleGlu	408
OY	1511	GGCTCCAAATCTGGGTCAATGAGGACCAACAGCTGTACTTCCAGGGCACCAAGAC	1570
Db	409	AlaLeu-----ThrSerAsp	413
OY	1571	ACGGCGTGGAGACACACCTCTACGTGTCAGCTATGAGCGCGCGC-----	1618
Db	414	-----TyrLeuTyrTyrIleSerAsnGlnTyrLysGlyMetProGlyGly	428
OY	1619	---GAGATCGTACGCTCACACAGCGCGGCTTCTCCATAGCTGTGCTCAATGACCCAGAC	1675
		:::	
Db	429	ArgAsnLeuTyrIleGlnIleLeuSerAspTyrThrLysValThrCysLeuSerCysGlu	448
OY	1676	TTGCACATGTTCTGTCAGCCACTACAGACAGCTGAGACAGCGCGCTGCGTCACTTAC	1735
		:::	
Db	449	LeuAsnProGlnArgCysGlnTyrTyrSerValSerPheSerLysGlnAlaLysTyrTyr	468
OY	1736	AAGCTG-----ACGGCGCC-----	1750
		:::	
Db	469	GlnLeuArgCysSerGlyProGlyLeuProLeuTyrThrLeuHisSerSerValAsnAsp	488
OY	1751	-----GACGACGACCCCTGACAAAGACAGCCCGCTTGGGCT	1789
Db	489	LysGlyLeuArgValLeuGlnAspAsnSerAlaLeuAspLys-----	502
OY	1790	AGCATGATGAGAGCGACACTGCGCGGATATGTTCTCTCCAGAAATCTTCATTTTC	1849
Db	503	---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe	516

OY	1850	CACACGGCGGATGTCGGGGCTC---TTGGCATGATGTATCAACCCCGACGCTTTCAG	1906
Db	517	IIelleuansngltuthrlyspheptrtyrGlnMetIleLeuProPheHis---Pheasp	535
OY	1907	CCAGGAGAGAACGCCACCCCGCTCTTTGTATATGAGAGCCCCAGGTGCAGCTGGTG	1966
Db	536	LysserlystysrtyrProleuLeuAspValtyrAlaGlyProGysserGlnyAla	555
OY	1967	AATACCTCTCTCAAGGCATCAAGTACTTGGCGCTCAAC-----ACACTGGGCTCC	2017
Db	556	AsphtrValPhe-----ArgleuAsnTrpAlaThrTyrLeuValSer	569
OY	2018	CTGGGCTACGCGCGTGGTGTG---ATTGACGGCAGGAGGGCTCCTGTCAACGAGGGCTTCGG	2074
Db	570	ThrcGlnsnIleIleValIalaSerPheaspGlyArgGlySerGlyTyrGlnGlnyAspIys	589
OY	2075	TTCCAGAGGGGCGCTGAAAAACCAATAGGGCCAGGTGGAGATGAGAGCACCAGTGGAGGC	2134
Db	590	IleMethIslaIleAsnArgAlaGluGlyThrPheGlnValGlnIlaSpGlnIleGlnIla	609
OY	2135	CTG---CAGTTGCTGGCGCGAGAGTATGGCTTCATGCACCTAGCGGAGGTGGCATCAT	2191
Db	610	AlaArgGlnPhe-----SerlysmetGlyPheValAspAsnIlyAsrGlyIleAlaIleTrp	627
OY	2192	GGCTGGTCTTACGGGGGCTTCCTCTGCTCATAGGGGCTAATCCACAAAGCCCGAGGTGTC	2251
Db	628	GlyTrpserTyrglyGlyTyrValIthrSerMetValIleuGlySerGlySerGlyValPhe	647
OY	2252	AAGGTGGCGAATCGGGGCTGCCCGCGTCACCGTCTGATGGCTACGACACAGGGAGCACT	2311
Db	648	LysGlyseltyIleAlaValAlaProValSerAlaGlyTrpGlyTyrTyraSpserValtyrThr	667
OY	2312	GAGCGCTACATATGACGCTC-----CCTGAGAACACACACAGCGCTTACGAGCGGGCTCC	2365
Db	668	GlnArglymetGlyLeuProThrProGlnAspAsnLeuAspHisTyrArgAsnSerThr	687
OY	2366	GTTGGCCCTGCACGTGGAGAAAGCTGCCAATGAGCCCAACCGCTTGATATCTCTCAGAGC	2425
Db	688	ValMetSerArgAlaGlnAsnPhe-----LysGlnValGlyTyrLeuIleuIleHisGly	705
OY	2426	TTCTGAGCAAGAAACGTGCACTTTTCCACACAACTCTCGTCTCCCAATGATGATC	2485
Db	706	ThrIlaAspAsnAsnValIaHisPheGlnGlnIleSerIaIleValIaLeuValAsp	725
OY	2486	GCAGGGAACCTTATCACAGCTCCAGATCTACCCACAGAGACACAGATATTCGCTGCCCC	2545
Db	726	ValIyValAspPheGlnIalaMetIleTyrThrAspGlnAsnHisGlyIleAlaSerSer	745
OY	2546	GAGTCGGGGCGAGCACTATGAAGTCACAGTCTGCTGCTCCATCTTCTACAGAA	2593
Db	746	ThrIlaHisGlnHisIleTyrThrHisMetSerHisPheIleuGln	761
RESULT 6			
US-09-794-236-1			
: Sequence 1, Application US/09794236			
: Patent No. 6337069			
: GENERAL INFORMATION:			
: APPLICANT: Grouzmann, Eric			
: APPLICANT: Lacroix, Jean-Silvain			
: APPLICANT: Monod, Michel			
: TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis			
: FILE REFERENCE: 81985/276823			
: CURRENT APPLICATION NUMBER: US/09/794,236			
: CURRENT FILING DATE: 2001-02-28			
: NUMBER OF SEQ ID NOS: 4			
: SOFTWARE: PatentIn version 3.0			
: SEQ ID NO 1			
: LENGTH: 766			
: TYPE: PRF			
: ORGANISM: Homo sapiens			
: US-09-794-236-1			
Alignment Scores:			


```

Db 556 AspThrValPhe-----ArGleuSnrTrpAlaThrTyrLeuAlaSer 569
QY 2018 CTGGGGCTACGCCCGGTGTG---ATTGACGGCAGGGCTCCGTGACGGCGGCTCGG 2074
Db 570 ThGluAsnIleIleValAlaSerPheAspGlyArgGlySerGlyTyrGlnGlyAspLys 589
QY 2075 TTGCAAGGGGCGCCGTAAGAAACCAATGAGGCGAGTGCAGATCGAGCAGCAGGTGAGGCG 2134
Db 590 IleMetHisAlaIleAsnArgLeuGlyThrPheGlnValGlnAspGlnIleGluAla 609
QY 2135 CTG---CAGTTCCGTGGCCGAGAGTATGCGTTCATCGACTGAGCCGAGTTGCCATCAT 2191
Db 610 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTyr 627
QY 2192 GGGTGGCTACAGGGGCGCTTCCTCGCTCATGGGCGCTAATCCACAGGCCCGCATGTC 2251
Db 628 GATTrpSerTyrGlyGlyTyrValThrSerMetValLeuGlySerGlySerGlyValPhe 647
QY 2252 AAGGTGGCATTCGGGGGTGCCCGGTCCAGTGTGATGGCTTCAGCAGAGGGTACACT 2311
Db 648 LysCysGlyIleAlaValAlaProValSerArgTrpGlyTyrLysPserValTyrThr 667
QY 2312 GAGCGCTACATGCAGCTC-----CCTGAGACACACGACGAGCTATGAGCGGGTTCC 2365
Db 668 GlnArgTyrMetGlyLeuProThrProGlnAspAsnLysPheLysTyrArgAsnSerThr 687
QY 2366 GTGGCCCTGCAGCTGAGAGAGCTGCCAATGAGCCCAACCGCTTATCCCTGCACGGC 2425
Db 688 ValMetSerArgAlaGlnAsnPhe-----LysGlnValGlnTyrLeuIleHisGly 705
QY 2426 TTCCCTGACGAAACGTCGACTTTTTCACACAAACTCTCGTCCCACTGATCCGA 2485
Db 706 ThrAlaSerAspAsnValHisPheGlnGlnSerAlaGlnIleSerLysAlaLeuValAsp 725
QY 2486 GCAGGGAACCTTACCGAGCTCCAGATCTACCCCAAGAGACAGACATTCGCTGCC 2545
Db 726 ValGlyValAspPheGlnAlaMetLysTrpTyrThrAspGlnAspHisGlyIleAlaSer 745
QY 2546 GAGTCGGGCGGACGACTATGAAGTCAGCTGCTGCACTTTCAGAGAA 2593
Db 746 ThrAlaHisGlnHisIleTyrThrHisMetSerHisPheIleLysGln 761

RESULT 7
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schlossman, Stuart F.
; APPLICANT: Tanaka, Yoshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 555X
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/833,211
; FILING DATE: 06-FEB-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3

Alignment Scores:
Pred. No.: 3.61e-29 Length: 755
Score: 482.50 Matches: 194
Percent Similarity: 36.47% Conservative: 116
Best Local Similarity: 22.82% Mismatches: 321
Query Match: 10.06% Indels: 219
DB: Gaps: 33

US-09-976-674-4 (1-2617) x PCT-US93-07923-3 (1-755)
QY 287 AAGAAAGTCCGGAAGAGAGGCTCTGCTCTGCTGTAAGAGCAGATGTCATTTTC 346
Db 39 LysAsnThrTyrArgLeuLysLeuTyrSerLeuArgTrp----- 51
QY 347 CAGGCCAGGCCCCACCATGGGGTCTACTCTGGGAGAGAGAGCTGTCG----- 394
Db 52 ---IleSerAspHisGlyTyrLeuTyrLysGlnLysAsnIleLeuValPheAsnAla 70
QY 395 -----AGGAGCGGAAAGCCCTGGGGGCTCCGCGCATACC 430
Db 71 GlnTyrGlyAsnSerSerValPheLeuGlnAsnSerThrPheAspGlnPheGlyHisSer 90
QY 431 TCCATGACGACTCCAC---AGCGAGAGTGGCTTCCCTTCCAGGCGACACAGCCTC 487
Db 91 IleAsnSprySerIleSerProAspGlyGlnPheIleLeuLeuGlnTyrAsnTyrVal 110
QY 488 TTCCACTGTGGCGAGCGGCGGACAGACGCTTCATGCTGTCCTTATGAACCGCTGAA 547
Db 111 LysGlnTrpArgHisSerTyrThrAlaSerTyrAspLeuValAsnLysArgGln 130
QY 548 ATCAAGACCCAGTCTGAGGCGCCGGATGAGCCCAAAATCTCCCGCCAGCCCGCC 607
Db 131 LeuIleThrGlnGluArgIleProAsnAsnThrGlnTrpValThrTyrPserProValGly 150
QY 608 TTC---TTCCTCTCATCAATTAACAGCGACGCTGGGTGGCCCAATCGAGACAGCGGAG 664
Db 151 HisLysLeuAlaTyrValITrPAsnAsnAspIleTyrValLysIleGlnProAsnLeuPro 170
QY 665 GAGCGGCGCTGACCTTCTGCCACCAAGTTTATCCAAATGCTCTGATGACCCCAAGTCT 724
Db 171 SerTyrArgIleThrTrp-----ThrGlyLysGlnAspIleIleTyrAsn----- 185
QY 725 GCGGGTGTGGCCACCTTCGTCATACAGAGAAG---TTCAGCGCTTCATCGGTACTGG 781
Db 186 ---GlyIleThrAspTrpValTyrGlnGlnGlnValPheSerAlaTyrSerAlaLeuTyr 204
QY 782 TGGTGGCCCAAGCGCTCTGAGGAGTTTCAGAGGCGCTCAAGACGCTGCAATCTGTAT 841
Db 205 TrpSerProAsnGlyThrPhe-----LeuAlaTyr 214
QY 842 GAGCAAGTCGATGAGTCGAGGTGAGGTCATTCAC-----GTCCCTCTCTGCG 892
Db 215 AlaGlnPheAsnAspThrGlnValProLeuIleLysTyrSerPheTyrSerAspGlnSer 234
QY 893 CTAGAAAGAAAGAGAGCGACTGTATCGGTACCCCGAGCAGCAGCAAGATCCCAAG 952
Db 235 LeuGlnTyrProLysThrValArgValProTyrProLysAlaGlyAlaValAsnProThr 254

```

```

QY 953 ATTGCTTGAACCTGGCTGAGTTCACAGTCCAGACCCAGAGATCGTCGACCCAG 1012
   :::: ||| ||||| ||| ::::
Db 255 Val-----LysPhePheValValAsnThrAspSerLeuSerValThrAsnAlaThr 272
QY 1013 GAGAGAGAGCTGGTGCACCCCTTCAGCTCGCTGTTCCGAAAGTGAGATACATCGCCAGG 1072
   :::: ||| ||| ||| ::::
Db 273 SerIleGlnIleThrAlaProAlaSerMetLeuIleGlyAsp---HisTyrLeuCysAsp 291
QY 1073 GCCGGGTGGACCCGGGATGGCAATACGCTGGGCCATGTTCTGGACGCCGCCAGCAG 1132
   ||| ||| ::::
Db 292 ValThrPheAlaThrGlnGluArgIleSer-----LeuGln 303
QY 1133 TGGCTCCAGCTGTCCTCCCTCCCGGCCCTGTTTCATCCGAGACAGAGATGAGAG 1192
   ||||| ||| ::::
Db 304 Tripleu----- 305
QY 1193 CAGCGGCTAGCCTGTCGACAGCTGTCCCGAGATGTCCAGCCGTATGTGTG----- 1246
   ||| :::: ||| |||
Db 306-----ArgArgIleGlnAsnTyrSerValMetAsp 315
QY 1247-----TAGAGAGAGCTCACCAGCTCGATCAATGTTTCATGACATCTTATCCC 1297
   ||||| ||| ::::
Db 316 IleCysAspTyrAspGluSerSerGlyArgTyr----- 326
QY 1298 TTCCCCCAATCAGAGGAGGAGAGAGAGCTGTCTTCCGCGCCATGAATGACAGACC 1357
   |||
Db 327-----AsnCysLeuVal 330
QY 1358 GGCCTTCGCAATTGTACAAAGTACCCGCGTTTAAATCCAGAGGCTACGATGGAGT 1417
   ||||| |||
Db 331 AlaArgGlnHisIleGluMetSerThr-----GlyTyrPheVal 343
QY 1418 GAGCCCTTCACCCCGGGGAGATGAA----- 1444
   ||| ||| |||
Db 344 GlyArgPheArgProSerGluProHisPheThrLeuAspGlyAsnSerPheTyrIle 363
QY 1445-----TTAAGTGCGCCATTAAAGAGAG 1468
   ||||| |||
Db 364 IleSerAsnGluGluGlyTyrArgHisIleCysTyrPheGlnIleAspIleAspIle 383
QY 1469 ATTGCTTGACAGCGGTGAATGGAGGTTTGGGAGGACAGCGGCTCCAGATCGGGTC 1528
   :::: ||| ||| ||||| |||
Db 384 ThrPheIleThrIleGlyThrThrProIleValIleGlyIleGluAlaLeu----- 399
QY 1529 AATGAGAGACCAAGCTGTACTTCCAGGGCACCAGACAGCCGCTGGAGCACAC 1588
   ||| ||| |||
Db 400-----ThrSerAsp-----Tyr 403
QY 1589 CTCTACGTGTCAGCTATGAGCGCGCGC-----GAGATGCTAGCGCTC 1633
   ||||| ||||| ||| |||
Db 404 LeuTyrTyrIleSerAsnGlnTyrIleGlyMetProGlyGlyArgAsnLeuTyrIle 423
QY 1634 ACCAGCGCGGCTTCTCCCATAGCTGTCCATGAGCCGAACTTCGACATGCTGTAC 1693
   :::: ||| ||| |||
Db 424 GlnLeuSerAspTyrThrIleValThrCysLeuSerCysGluLeuAsnProGluArg 443
QY 1694 CACTACAGACAGCTGAGCAGCGCGCTGCGTCCAGCTGTACAGCTG-----AGCGGC 1747
   ||| ||||| ||||| |||
Db 444 GlnTyrTyrSerValSerPheSerIleGluAlaIleGlyTyrGlnLeuArgCysSer 463
QY 1748 CCC----- 1750
   |||
Db 464 ProGlyLeuProLeuTyrThrLeuHisSerSerValAsnAspIleGlyLeuArgVal 483
QY 1751---GAGCAGCAGCCCTGACAGACAGCCCGCTTCGCGGTAGCATGATGAGGACAGCC 1807
   ||||| ||| |||
Db 484 GluAspAsnSerAlaLeuAspIle-----MetLeuGlnAsnVal 496
QY 1808 AGCTGCCCGCGGATTAGTTCCTCCAGAGATCTTCCATTTCACAGCCGCTGGATGTG 1867
   :::: ||| ||| |||
Db 497 Gln-----MetProSerIleTyrIleAspPheIleIleLeuAsnGlnThr 511

```

```

QY 1868 CGGCTC---TAGCGCATGATCTACAGCCCGCCCTTCGACCCAGGAGAGAACCC 1924
   :::: ||| ||||| |||||
Db 512 LysPheTyrPyrGlnMetIleLeuProProHis---PheAspIleSerIleTyrPro 530
QY 1925 ACCGCTCTTTGTATATGAGGCCCGCCAGGTGACGTGTGAATTAATCTCTCAAGGC 1984
   :::: ||||| ||||| |||
Db 531 LeuLeuAspValTyrAlaGlyProCysSerGlnIleAlaAspThrValPhe----- 548
QY 1985 ATCAAGTACTTGGCGCTAAC-----ACATGGCGCTCCCGGGCTACCGCGTGT 2035
   ||||| ||||| |||||
Db 549-----ArgLeuAsnThrPheAlaThrTyrIleLeuAlaSerThrGluAsnIleVal 564
QY 2036 GTG---ATTGACGCGAGGCGCTCTGTCAGGAGCGCTTCGTTCAAGAGCGCCCTGAA 2092
   ||||| ||||| |||||
Db 565 AlaSerPheAspGlyArgGlySerGlyTyrGlnGlyAspIleMetHisAlaIleAsn 584
QY 2093 AACCAATGGCCAGGTGAGATGAGAGACAGCAGGTGAGGCGCTG---CAATTCGTGGC 2149
   :::: ||| ||||| |||||
Db 585 ArgArgLeuGlyThrPheGluValGluAspGlnIleGluAlaIleArgGlnPhe----- 602
QY 2150 GAGAGATGGCTTCATGACCTGACCTGACCGAGTTCATTCATGCTGCTCAGGGGCG 2209
   ||| ||||| ||||| |||||
Db 603 SerIleMetGlyPheValAspAsnIleAlaIleTyrGlyTyrPheSerTyrGly 622
QY 2210 TTCCCTGCTCATGAGGCGCTATGCCACAGCCCGAGTTCGATCAAGGTGCGCGGT 2269
   :::: ||| ||| |||||
Db 623 TyrValThrSerMetValLeuGlySerGlyValIlePheIleValIleVal 642
QY 2270 GCGCCGCTACCGCTGCTGATGAGCTGACAGCAGGCTACCTGACGTGACGAGCTG 2329
   ||||| ||||| ||||| |||||
Db 643 AlaProValSerArgTyrGlyTyrAspSerValTyrGlnGlyArgTyrMetGlyLeu 662
QY 2330-----CCTGAGACAAACACAGCAGCGCTATGAGCGGGTTCGTCGCTGACAGT 2383
   ||||| ||||| |||
Db 663 ProThrProGluAspAsnLeuAspHisTyrArgAsnSerThrValMetSerArgAla 682
QY 2384 AAGCTGCCAATGAGCCCAACCGCTGTCTATCCGACCGCTTCGACGAGAAACGTG 2443
   :::: ||| ||||| |||||
Db 683 AsnPhe-----LysGlnValGluTyrLeuLeuIleHisGlyThrAlaAspAsnVal 700
QY 2444 CACTTTTTCACACAAACTCTCTGCTCCCAACTATGACGAGCAAGAACTTACAG 2503
   ||||| ||| :::: |||
Db 701 HisPheGlnGlnSerAlaGlnIleSerIleAlaLeuValAspValGlyValAspPheGln 720
QY 2504 CTCGAGATCTACCCCAACGAGAGACAGATATGCTGCTCCCGAGTCCGGGAGACATAT 2563
   ||| ||| ||| |||
Db 721 AlaMetTyrTyrThrAspGluAspHisGlyIleAlaSerSerThrAlaHisGlnHisIle 740
QY 2564 GAAGTCAGCTGCTGCTACTTCTACAGGAA 2593
   :::: ||||| |||||
Db 741 TyrThrHisMetSerHisPheIleIleGln 750

RESULT 8
US-09-462-284-2
: Sequence 2, Application US/09462284
: Patent No. 6309868
: GENERAL INFORMATION:
: APPLICANT: Nestec S.A.
: APPLICANT: Monod, Michel
: APPLICANT: Doumas, Agnes
: APPLICANT: Affolter, Michael
: APPLICANT: Van den Broek, Peter
: TITLE OF INVENTION: CLONING OF THE
: TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
: FILE REFERENCE: 8265-298
: CURRENT APPLICATION NUMBER: US/09/462,284
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 771
: TYPE: PRT

```


Db 717 ProGluLysLeuHisSer 722
||| :|||
RESULT 9
US-08-230-491A-2
Sequence 2, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Retlig, Wolfgang J.; Scanlan, Matthew J.;
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELFE & LYNCH
STREET: 805 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-2

Alignment Scores:
Pred. No.: 8.64e-25 Length: 760
Score: 426.00 Matches: 168
Percent Similarity: 38.85% Conservative: 109
Best Local Similarity: 23.56% Mismatches: 256
Query Match: 8.88% Indels: 180
Gaps: 31
US-09-976-674-4 (1-2617) x US-08-230-491A-2 (1-760)
QY 593 CCTCCGACCGCTGCTTCTTCTTCATCAATAAGACGACGTGGTGCCCAATC 652
||| ||||| |||
Db 176 ProGluLysProProPhe----- 181
QY 653 GAGACAGCGGAGGAGCGGCTGACCTTCGCCACCAAGTTTATCCATGCTCGAT 712
||| :||| |||
Db 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
QY 713 GACCCCAAGTGTGGGGTGTGGCCACTTCTCATACGAGGAAGGTTGACCGCTTCACT 772
||| :||| |||
Db 194 Asn-----GlyIleProAspTyrPalaTyrGluGluMet--LeuProThr 208
||| :||| |||
QY 773 GGGTAC-----TGTGGTGGCCCAACACCTCTGGAAGGTTCAAGGCGCTCAAGAG 826
||| ||||| |||
Db 209 LysTyrIleLeuTyrPripSerProAsnGlyLysPhe----- 220
QY 827 CTGGCAATCTGTATGAGGAAGTGTGAGTGTGAGGTGAGGTCAATTCAGTCCCTCT 886
||| ||| :||| |||
Db 221 -----LeuAlaTyrAlaGluPheAsnAspLysIleProValIleAlaTyrSerTyr 238

QY 887 CCTGGCGCTAGAGAA---AGGAAGCGGACTGCTATCGTACCCGACAGCAGCAGAG 943
||| :||| :||| :|||
Db 239 TyrGluLysPrgLugInLysProArgThrIleAsnIleProTyrProLysAlaGluLys 258
QY 944 AATCCCAAGATTGCTTGAACCTGCTGAGTTCACAGACTGACAGCCAGCGCAATCGTC 1003
||| ||| :||| :|||
Db 259 AsnProValAlaArgIle-----PheIleIle 267
QY 1004 TCGACCCAGGAGAAAGAGACTGCTGACGCCCTTCAGCTCGCTTCCGAGAGTG- 1057
||| ||| :||| |||
Db 268 AspThrThrTyrProAlaTyrValGlyProGluGluValProValProAlaMetIleAla 287
QY 1058 -----GAGTCAATCGCCAGGGCGCGGTGACAGCCGGATGGCAATATAGCCTGGCC 1108
||| :||| :||| :|||
Db 288 SerSerAspTyrTyrPheSerThrPheThrPalaThrAspGluArgValCys----- 305
QY 1109 ATGTTCCTGGACCGCCAGCCAGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1168
||| ||||| :||| |||
Db 306 -----LeuGlnThrPheLysArgValGlnAsnValSerValLeuSer 319
QY 1169 ATC-----CCGACGACAGAGAT----- 1186
||| :||| :||| :|||
Db 320 IleCysAspPheArgGluAspTyrPglInThrTrpAspCysProLysThrGluHisIle 339
QY 1187 GAGAGAGCGGGCTA-----GCCTTCCGACAGCTGCTCCAGAGATGTCACCGCTAT 1240
||| ||| ||| |||
Db 340 GluGluSerArgThrGlyTyrPalaGlyLysPhePheValSerAlaProValPheSerTyr 359
QY 1241 GTGTGTACGAGAGGATCACCAGCTGTGATCATGTATGACATCTTATCCCTTC 1300
||| :||| :||| :|||
Db 359 ----- 359
QY 1301 CCCCAATCAGAGGAGAGAGAGAGCTGCTTCTCCGCCAATGATGCAAGCCGCGC 1360
||| :||| :||| :|||
Db 360 -----AspAlaIleSerTyrLysIlePheSerAspLysAspLys 373
QY 1361 TTCGCACTTGTACAAAGTACCGCGCTTTAAATCCAGGGCTACGATGGAGTCAG 1420
||| :||| :||| :|||
Db 374 TyrLysHisIleHisTyrIle----- 380
QY 1421 CCTTCAGCCCGGGGAGATGATTAAGTCCCATTAAGGAAGATGCTGTGACC 1480
||| :||| :||| :|||
Db 381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
QY 1481 AGCGGTGAATGGAGGTTTGGCGAGCGACGCTCCAAAGATCTGGTCAATGAGAGACC 1540
||| :||| :||| :|||
Db 392 SerGlyLysTyrPglAlaIle-----AsnIlePheArgValThrGluAsp 406
QY 1541 AAGCTGGTACTCTCCAGGCGAC--AAGACACGCGCGTGGAGCACCACCTCTAC-- 1594
||| ||| :||| |||
Db 407 SerLeuPheTyrSerSerAsnGluPheGluGluTyrProGlyArgArgAsnIleTyrArg 426
QY 1595 -----GTGGTACGCTATGAGCGCGCGAGATGAGTCAAGCTCAGACAGCCGCGCTTC 1648
||| :||| :||| :|||
Db 427 IleSerIleGlySerTyrProProSerLysLysCysVal----- 439
QY 1649 TCCCATAGCTGCTCCATG-----AGCCAGAACTTGAGCATG 1684
||| :||| :||| :|||
Db 440 -----ThrCysHisLeuArgLysGluArgCysGlnTyrThrAlaSerPheSerAsp 457
QY 1685 TTCGTACGCCATACAGAGGCTGAGCAGCGCGCGCTCGTCCAGCTGTACAGCTG-- 1741
||| :||| :||| :|||
Db 458 TyrAlaTyrTyrAlaLeuValCysTyrGlyProGlyIleProIleSerThrLeuHis 477
QY 1742 AGCGGCCCGGAGAGACCCCTGCAC--AAGCAGCCCGCTTCTGGGTAGCATG 1795
||| ||| :||| :|||
Db 478 AspGlyArgThrAspGlnGluIleLysIleLeuGluGluAsnLysGluLeuGluAsnAla 497
QY 1796 ATGAGCAGCAGCAGCTGCGCCCGGATATGCTTCACAGATCTTCATTTCCACAGC 1855
||| ||| :||| :|||
Db 498 LeuLysAsnIleGlnLeuProLysGlu-----GluIleLysLysLeuGluVal 513


```

QY 1856 CGCTGGAGTGGCGGCTTACAGCATGATCTACAGGCCAGCGCTTGACAGGAG 1915
      ||| ||||| |||
Db 514 AspGluIleThrLeuTrpTyrLysMetIleLeuProGln---PheAspArgSerLys 532
      ||| ||||| |||
QY 1916 AAGCACCACCGCTCCCTTTATATGAGGCCCGCAGGTGACGTGGATTAATCTC 1975
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 533 LysTyrProLeuLeuIleGlnValTyrGlyGlyProCysSerGlnSerValArgSerVal 552
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1976 TTC-----AAGGACATGACTCTTCCGGCTCAACACACTGGCTCCCTGGCTACGCC 2029
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGluGlyMetVal 568
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2030 GTGGTTGTGATGACGCGAGGGGCTCTGTCAGGAGGGCTTGCTGGAAGGGCCCTG 2089
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnIleLysPheLeuTyrAlaVal 588
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2090 AAAACCAATGGCGCCGAGGATGAGACCGACAGTGGAGGCCCTGCAGTTCGTGCC 2149
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 589 TyrArgLysLeuGlyValTyrGluValGluAspIleThrAlaValArgLysPheIle 608
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2150 GAGAGATGAGCTTCATGCACTGAGCGGAGTTGCAATCCATGGCTGCTACGGGGG 2209
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 609 Glu---MetLysPheIleAspGlyLysArgIleAlaIleTrpGlyTyrPserTyrGlyGly 627
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2210 TTCCTGCTGCTAGGGGCTAATCCAGACGCCCGAGGTTCAGAGTGGCATTCGCGGT 2269
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 TyrValSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysCysGlyIleAlaVal 647
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2270 GCGCGGCTGACGCTGATGGGCTAGACAGAGGTAAGCGGCTGACATGAGAGCTC 2329
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 AlaProValSerSerTrpGlyTyrTyrAlaSerValTyrThrGluAlaArgPheMetGlyLeu 667
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2330 CCT-----GAGAACACAGACAGCGGCTATAGCGGGGCTCCGTCGTCAGAGTGGAG 2383
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 668 ProThrLysAspAsnLeuGlnIleLysTyrLysAsnSerThrValMetAlaArgAlaGlu 687
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2384 AAGTCCCAATGAGCCCAACCGCTTATCTCTCCAGCGCTTCTGGACGAAACGTG 2443
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAspAsnVal 705
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2444 CACTTTTCCACCAAACTCTGCTCCCACTGATCCGAGCGAGGGAACCTTACAG 2503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 706 HisPheIleAsnSerAlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGln 725
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2504 CTCGAGATCTACCCCAAGAGACAGATATTCGTCGCCCGAGGCGGC----- 2554
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 726 AlaMetTrpTyrSerArgGlnAsnHisGlyLeu-----SerGlyLeuSerThr 741
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2555 GAGCACTAGAAGTCAGCTTGCTGCACTTCTTACAGGAA 2593
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 742 AsnHisLeuTyrThrHisMetThrHisPheLeuGln 754
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

      ; SOFTWARE: WordPerfect
      ; CURRENT APPLICATION DATA:
      ; APPLICATION NUMBER: US/08/619,280A
      ; FILING DATE: 18-MARCH-1996
      ; CLASSIFICATION: 435
      ; PRIOR APPLICATION DATA:
      ; APPLICATION NUMBER: 08/230,491
      ; FILING DATE: 20-APRIL-1994
      ; ATTORNEY/AGENT INFORMATION:
      ; NAME: Hanson, No. 5767242man D.
      ; REGISTRATION NUMBER: 30,946
      ; REFERENCE/DOCKET NUMBER: LUD 5350.1
      ; TELECOMMUNICATION INFORMATION:
      ; TELEPHONE: (212) 688-9200
      ; TELEFAX: (212) 838-3884
      ; INFORMATION FOR SEQ. ID NO. 2:
      ; SEQUENCE CHARACTERISTICS:
      ; LENGTH: 760 amino acids
      ; TYPE: amino acid
      ; TOPOLOGY: linear
      ;
US-08-619-280A-2
      ;
Alignment Scores:
      Pred. No.: 8,64e-25 Length: 760
      Score: 426.00 Matches: 168
      Percent Similarity: 38.85% Conservative: 109
      Best Local Similarity: 23.56% Mismatches: 256
      Query Match: 8.88% Indels: 180
      DB: 1 Gaps: 31
      ;
US-09-976-674-4 (1-2617) x US-08-619-280A-2 (1-760)
QY 593 CTGCGGACCGCTCTCTTCTCTTATCATATACAGCAGCCTGTGGTCCCAACATC 652
      ||| ||||| |||
Db 176 ProGlyAspProProPhe----- 181
QY 653 GAGACAGGCGAGAGCGCGGCTGACCTTGCCACCAAGTTTATCAATGTCGTGAT 712
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 713 GACCCCAAGTCTGCGGCTGGCGCACCTTGTCATACAGAGAGAGTTCGACCCCTCACT 772
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194 Asn-----GlyIleProAspTrpValTyrGluGluIleMet---LeuProThr 208
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 773 GGGTAC-----TGGTGGTCCCGCCACAGCCTCTGGGAAAGTTTCAGAGGCTCAAGAGC 826
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 LysTyrAlaLeuTrpTrpSerProAsnGlyLysPhe----- 220
QY 827 CTGCGAATCTGTATGAGAAATCGATGATCCGAGGTGAGGTCAATTCACGTCCCTCT 886
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 -----LeuAlaTyrAlaGluPheAsnAspLysAspIleProValIleAlaTyrSerTyr 238
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 CTGCGCTAAGAA---AGAAAGAGGAGTGGATTCGATCCGACCCAGGACAGCAGAG 943
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TyrGlyAspGluGlnTyrProAlaGlyThrIleAsnIleProTyrProLysAlaGlyLys 258
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 944 AATCCCAAGATTCCTGAAATGCGTGAAGTCCAGATCAGACAGCGCAAGATCGTC 1003
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 AspProValAlaArgIle-----PheIleIle 267
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1004 TCGACCCAGAGAGAGAGCTGTCAGCCCTTCAGCTCGCTGTTCCCGAAGTG----- 1057
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 AspThrThrTyrProAlaTyrValGlyProGlnGluValProValProAlaMetIleAla 287
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1058 -----GAGTACATCGCGAGGCGCGGTGGACCGCGAGTGCACAAATAGCGCTGGGCC 1108
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 SerSerAspTyrTyrPheSerTrpLeuThrTrpValThrAspAlaArgLys----- 305
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1109 ATGTTCTGGACCGCGCCAGAGTGGCTCAGCTGCTCTCCGCCCGCGCTGTTTC 1168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 306 -----LeuGlnTrpLeuLysArgValGlnAsnValSerValLeuSer 319
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1169 ATC-----CCGAGCACAGAGAT----- 1186

```


Db	320	Illecysaphehrvgluasrptpqlnthrtpraspocyprolyvthcgluinhstle	339
Oy	1187	GAGAGACAGGGGCTA-----GCCTTGCCAGAGCTGCCCCAGAAATGTCACCGCTAT	1240
Db	340	Gluluseratrgthrglytrpalaiglygphhevalserargprovalphesertry	359
Oy	1241	GTGGTGTACAGAGAGGTCCACCAACGTCTGGATCAATGTCATGACATCTCTATCCCTTC	1300
Db	359	-----	359
Oy	1301	CCCCAATCAGAGAGAGAGACAGACAGCTGCTCTTCGCCCAATGAATGCAGACCGGC	1360
Db	360	-----AspAlaIleSerTyrTyrLysIlePheSerAspLysaspoly	373
Oy	1361	TTTCGCATTGTGTACAAATGCACCGCGCTTTTAAATCCAGAGGCTACGATTGAGTAG	1420
Db	374	TyrLysHstIleHstYrIle-----	380
Oy	1421	CCCTTCACGCCCGGGAAGATGAATTTAATGCCCCCATTAAGAAAGATTCCTGTGAC	1480
Db	381	-----LysAspPthValGluAsnAlaIleGlnIleHthr	391
Oy	1481	AGCGGTAAATGGAGAGTGTGTGGCGAGGACAGCGCTCCAAATGTGGGTGTCATAGAGAGAC	1540
Db	392	SerGlyLysrtpgluaIle-----AsnIlePheargYalThrGlnsr	406
Oy	1541	AASTGGTGTACTTCCAGGGCACCC---AAGACACGCGCGCTGAGACACACCTCTAC	1594
Db	407	SerLeupheYrSerSerAsnGluPheGluGluTyrProGlyArgArgsnIleYrarg	426
Oy	1595	-----GTGGTCACACTGTAGAGGGGGCGCGAGATGTGACGCTCACACCGCCGGCTTC	1648
Db	427	IleSerIleGlySerProProSerLysLysCysVal-----	439
Oy	1649	TCCCATGAGCTGCTCCATG-----AGCAGAACTTCGACATG	1684
Db	440	-----ThrCysHsLeuAlaGlyGluAlaGlyCysGlnTyrTyrThrAlaSerPheSerasp	457
Oy	1685	TTTCGTGACCCACTACAGACAGGTGAGACGCGCGCTGCGTACAGCTGACAAAGTGTG---	1741
Db	458	TyrAlaLysTyrTyrAlaLeuValCysTyrGlyProGlyIleProIleSerThrIleuHs	477
Oy	1742	AGCGGCCCGCCAGACGACCCCTGTGCAC-----AAGCAGCCCGCTTCTGGGCTAGCATG	1795
Db	478	AspGlyArgThrAspGlnGluIleLysIleLeuGlnGluAsnLysGluLeuGluAsnAla	497
Oy	1796	ATGAGGAGGACGACAGCTGCCCGCGAATTAATGTTCCCTCCAGACATCTTCATTTCCACAG	1855
Db	498	LeuLysAsnIleGlnLeuProLysGlu-----GluIleLysLysLeuVal	513
Oy	1856	CGCTCGATGTGGCGCTCTACGCGCATGATCAACAGCCCGAGCTTCGAGCAGGAGAG	1915
Db	514	AspGluIleThrIleuTyrTyrLysMetIleLeuProProGln---PheAspArgSerLys	532
Oy	1916	AAGCAGCCCACTGCTCTTGTATATGAGAGCGCCCGACAGTGCAGCTGTGAATATCC	1975
Db	533	LysTyrProLeuIleGlnValTyrGlyGlyProCysSerGlnSerValArgSerVal	552
Oy	1976	TTC-----AAAGCATCAAGTACTTGGGGCTCAACACACTGGCTCCCTGGGCTAGCC	2022
Db	553	PheAlaValAsnTrrIleSerTyrLeu-----AlaSerLysGlnLysIleVal	568
Oy	2030	GTGCTGTGATGATGAGCGAGAGGCTGCTGTCAGACGAGGCGTTCGGTTCAAGAGCGCTG	2089
Db	569	IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyAspLysIleuLeuTyrAlaVal	588
Oy	2090	AAAAACAATGGCGCAGGTGAGATCGAGACAGCAGGTGAGAGGCCCTGCAGTTCTGTGCC	2149
Db	589	TyrArgLysLeuGlyValTyrGluValGluAsnGlnIleThrAlaValArgLysPheIle	608
Oy	2150	GAGAAATGTGGCTTCATGACCTGACCGCCAGATGGCTGCATCATGCTGGTCTCAGGGGCG	2209

```

Db      609  glu---metglrphelleasrgilulysargilalealallerprdllytrpserlytglngly  627
QY      2210  TTQCTGTGGTCATGGGGGTAAATGCACAGAGCCCGAGGTGTCAAGGTGGCCATCGCGGGT  2265
Db      628  ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  647
QY      2270  GCGCGGCGACCGCGTCTGGATGGCGCTACGACACAGAGGTACACACTGAGCGGTACATGACATGC  2322
Db      648  AlaProValSerSerTyrGluTyrTyrTyrAlaSerValTyrTyrGlnIlyrGlnrPheMetGlyLeu  667
QY      2330  CCT-----GAGAACACACAGAGCGAGGTGTGAGAGCGCGGGTCCCTGGCGCCCTGCACGTGAG  2388
Db      668  ProThrIysAspAspAsnLeuGlnIleHisTyrLysAsnSerThrValIleMetAlaArgAlu  687
QY      2384  AAGTGGCCCAATGAGCCCAACCGGTGCTTATCGTCCACGCGCTTCTGTGAGAGAAAGCTG  2443
Db      688  TyrPheArgAsnValAsp-----TyrLeuLeuIleHisGlyThrAlaAspAsnVal  705
QY      2444  CAGTTTTTCCACACAAACTTCGTGTCTGCCAACTATTCGAGACAGAGAAACSTTACAG  2500
Db      706  HisPheGlnAsnSerIleAlaGlnIleAlaLysAlaLeuValAsnAlaGlnValAspPheGln  725
QY      2504  CTCGAGATCTGCCACCAAGACAGACAGATAGTTCGTCGCCCGCAGTGGCGGC-----  2555
Db      726  AlaMetTrpTyrSerIAspGlnAsnHisGlyLeu-----SerGlyLeuSerThr  741
QY      2555  GAGCAGTATGAAGTCAGCTGTGGTGCACATTCTACAGGAA  2593
Db      742  AsnHisLeuTyrThrHisMetThrHisPheLeuLysGln  754

RESULT 11
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,391
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5965373man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;

```

US-08-940-391-2

Alignment Scores:

Pred. No.:	8,64e-25	Length:	760
Score:	426.00	Matches:	168
Percent Similarity:	38.85%	Conservative:	109
Best Local Similarity:	23.56%	Mismatches:	256
Query Match:	8.88%	Indels:	180
DB:	2	Gaps:	31

US-09-976-674-4 (1-2617) x US-08-940-391-2 (1-760)

```

QY 593 CCGGCGACCCCTGCTTCTCTCTCATATACAGCAGCTGGTGGCCACATC 652
    ||| ||||| |||
Db 176 ProGlyAspProPhe-----
QY 653 GAGACAGGCGAGGCGGCGTGCACCTTGCACAGGTTTATCCATGTCGGAT 712
    ||| ||||| |||
Db 182 -----GlnIleThrPhe-----AsnGlyArgGluAsnLysIlePhe 193
QY 713 GACCCCAAGTCTGGGGGTGGCCACCTTGCATACAGAGAGTTCAGCCGCTTCACT 772
    ||| ||||| |||
Db 194 Asn-----GlyIleProAspTrpValTyrGluGluMet-----LeuProThr 208
QY 773 GGGTAC-----TGGTGGTCCCCACAGCCTCTGGGAGGTTTCAGAGCCTCAGACG 826
    ||| ||||| |||
Db 209 LysTyrIleLeuTrpTrpSerProAsnGlyLysPhe-----
QY 827 CTCGGAATCTGTATGAGAGTGCATGCGAGGTGAGGTCAGTTCACGTCCCTCT 886
    ||| ||||| |||
Db 221 -----LeuAlaTyrAlaGluPheAsnAspLysSpIleProValIleAlaTyrSerTyr 238
QY 887 CCGTGGCTAGAGAA--AGGAAGACGAGCTCGATGTCATCCCAAGAGCAGCAG 943
    ||| ||||| |||
Db 239 TyrGlyAspGluGlnTyrProArgThrIleAsnIleProTyrProLysAlaGlyAlaLys 258
QY 944 AATCCCAAGATTGCTTGAATAGTGGTGAAGTCCAGACTCAGACTCAGACAGCAGGCAAGATGTC 1003
    ||| ||||| |||
Db 259 AsnProValAlaArgIle-----PheIleIle 267
QY 1004 TCGACCCAGAGAGAGAGTGTGTCAGCCCTTGCAGTCCGCTGTTCCGAGAGTG- 1057
    ||| ||||| |||
Db 268 AspThrThrTyrProAlaTyrValGlyProGluValProValProAlaMetIleAla 287
QY 1058 -----GAGTACATCGCCAGGCGCGGTGAGACCGGAGTGAATACGCTGGGCC 1108
    ||| ||||| |||
Db 288 SerSerAspTyrTyrPheSerTrpLeuThrTrpValThrAspGluArgValCys----- 305
QY 1109 ATGTCTGTGACCGGCCCCAGAGTGGCTCAGCTGCTCTCTCCCGCGCCCTGTTC 1168
    ||| ||||| |||
Db 306 -----LeuGlnTrpLeuLysArgValGlnAsnValSerValIleuSer 319
QY 1169 ATC-----CCGACACAGAGAT----- 1186
    ||| ||||| |||
Db 320 IleCysAspPheArgGluAspTrpGlnThrTrpAspCysTrpLysThrGlnGlnHisIle 339
QY 1187 GAGGACAGCGGCTA-----GCCCTGCCAGAGCTGTCCCGAGGAATGTCCAGCCGAT 1240
    ||| ||||| |||
Db 340 GluGlnSerArgThrGlyTrpAlaGlyGlyPhePheValSerArgProValPheSerTyr 359
QY 1241 GTGTGTAGAGAGAGTCCACCAACGTGTGATCAATGTTATGACATCTTCTATCCCTTC 1300
    ||| ||||| |||
Db 359 ----- 359
QY 1301 CCCCAATCAGAGGAGAGAGAGCTGTCTTCTCCGCCCAATGATGACAGACCGGC 1360
    ||| ||||| |||
Db 360 -----AspAlaIleSerTyrTyrLysIlePheSerAspLysAspGly 373
QY 1361 TTCTGCATTGTGACAAAGTACCGCGCTTTAAATCCAGGCGTACGATTGGAGTGA 1420
    ||| ||||| |||
Db 374 TyrLysHisIleHisTyrIle----- 380
QY 1421 CCCTTACGCCCCGGGAGAGATGAATTAAGTCCCATTTAAGAGAGATGCTGTGACC 1480

```

```

Db 381 -----LysAspThrValGluAsnAlaIleGlnIleThr 391
QY 1481 AGCGGTGAATGGAGGTTTGGGAGAGCAGCGGCTCCAAATCTGGGTCATGAGGAGACC 1540
    ||| ||||| |||||
Db 392 SerGlyLysTrpGluAlaIle-----AsnIlePheArgValThrGlnAsp 406
QY 1541 AAGCTGTGTACTTCCAGGGCACCC--AAGACAGCGCGCTGGAGACACACACCTAC-- 1594
    ||| ||||| |||
Db 407 SerLeuPheTyrSerSerAsnGluPheGluGluTyrProGlyArgAsnIleTyrArg 426
QY 1595 -----GTGGTCACTATGAGCGGCGCGGAGATGCTACGCGCTCAGACCGCGGCTTC 1648
    ||| ||||| |||
Db 427 IleSerIleGlySerTyrProProSerLysCysVal----- 439
QY 1649 TCCCATGACTCTCCATC-----AGCCAGACTTCGACATG 1684
    ||| ||||| |||
Db 440 -----ThrCysHisLeuAlaGlyGluAlaGlyGlnIleTyrThrAlaSerPheSerAsp 457
QY 1685 TTGTCAGCCCATACAGACAGTGCAGACCGCGCGCTGGTGCACGTTCAAGCTG-- 1741
    ||| ||||| |||
Db 458 TyrAlaLysTyrGlyAlaLeuValCysTyrGlyProGlyIleProIleSerThrIleHis 477
QY 1742 AGCGGCGCGACAGACAGCCCTGAC- --AAGACCGCGCTTCTGGCTAGCATG 1795
    ||| ||||| |||
Db 478 AspGlyArgThrAspGlnGluIleLysIleLeuGluGlnLysGluLeuGluAsnAla 497
QY 1796 ATGAGGACACACAGCTGCCCGGATTAATGTTCTCCAGATCTCCATTTCCACAGC 1855
    ||| ||||| |||
Db 498 LeuLysAsnIleGlnLeuProLysGlu-----GluIleLysLysLeuGluVal 513
QY 1856 CGCTCGATGTGGCGCTTACAGGATGATCTACAGGCCACCGCCTTGCAGCCAGGAG 1915
    ||| ||||| |||
Db 514 AspGlnIleThrLeuTrpTyrLysMetIleLeuProGln--PheAspArgSerLys 532
QY 1916 AAGCACCCACCGCTCTTGTATATGAGGCGCCAGGCTGCAGCTGGAATTAAGTCC 1975
    ||| ||||| |||
Db 533 LysTyrProLeuLeuIleGlnValTyrGlyLysProCysSerIleSerValArgSerVal 552
QY 1976 TTC-----AAAGCATCAAGTACTTGGCGCTCAGACACACTGGCCTCCCTGAGGCTACGCC 2029
    ||| ||||| |||
Db 553 PheAlaValAsnTrpIleSerTyrLeu-----AlaSerLysGluGluMetVal 568
QY 2030 GTGGTGTGATTAAGCGAGGCGCTCTGTACAGAGGCTTGGTTCGAAGGCGCTG 2089
    ||| ||||| |||
Db 569 IleAlaLeuValAspGlyArgGlyThrAlaPheGlnGlyAspLysLeuLeuTyrAlaVal 588
QY 2090 AAAAACCAATGGCCAGGTGAGATCGAGACACAGGTCAGAGGCGCTGCAAGTTCGTGGCC 2149
    ||| ||||| |||
Db 589 TyrArgLysLeuGlyValTyrGluValGluAspGlnIleThrAlaValArgLysPheIle 608
QY 2150 GAGAAGTATGCTTTCATCGACCTGAGCGGAGTGCATCCATGATGCTAGCGGCGC 2209
    ||| ||||| |||
Db 609 Glu---MetGlyPheIleAspGluLysArgIleAlaIleTrpGlyTrpSerTyrGlyGly 627
QY 2210 TTCTCTGCTCATGGGGCTAATTCACAAAGCCCAAGGTTCAGAGTGCATCCGCGGT 2269
    ||| ||||| |||
Db 628 TyrValSerSerLeuAlaLeuAlaSerGlyThrGlyLeuPheLysGlyIleAlaVal 647
QY 2270 GCCCGGTACCGCTGAGTGGCTTACGACAGAGGATACAGGCTACAGGCTACAGAGCTC 2329
    ||| ||||| |||
Db 648 AlaProValSerSerTrpGlyTyrAlaSerValTyrThrGluArgPheMetGlyLeu 667
QY 2330 CCT-----GAGAACAACAGCAGCAGGCTATGAGCGGGTTCCTGTGCGCTGACGTTGAG 2383
    ||| ||||| |||
Db 668 ProThrLysAspAsnLeuGlnHisTyrLysAsnSerThrValMetAlaArgAlaGlu 687
QY 2384 AAGCTGCCAATGAGCCCAACCGCTTGCATTCACAGCGCTTCTGACGAAACGTTG 2443
    ||| ||||| |||
Db 688 TyrPheArgAsnValAsp-----TyrLeuLeuIleHisIleThrAlaAspAsnVal 705
QY 2444 CACTTTTCCACAACAACTTCTCGTCTCCCAACTGATCGAGACAGGAAACCTTACAG 2503
    ||||| ||||| |||

```



```

QY 1112 ACATGCCCCAGGCGATTATTGCGATCCCGGGTCCACCGGCCCTGGCATGTACTCCACCT 1053
Db 403 -----ProGlySerArgAspPro-----GlyProPro 411
QY 1052 TCGGGAACAGCGACTGTAAGGGCTGCACACAGCTCTTCTCTGGGTGCAGACGATCTTGC 993
Db 412 G1yAla-----ProGlyProAlaGlyProProGlySerArgAspProGly 426
QY 992 CCTGGCTGTGAGTGGAACT-----CAGCCAGTTTCAAGCAATCTTGGGATTTCTGC 939
Db 427 ProProGlyAlaProGlyProAlaGlyProProGlySerArgAspPro----- 442
QY 938 TGCCTGTCTGGGGTACCAGATACAGATCCGCTTCTCTTCTTACGGCAGAGAGGGGA 879
Db 443 -----GlyProProGlyAlaProGlyPro-----AlaGlyProProGlySer 456
QY 878 CGTGAATGACCTCCACTCCGAGCTCAT-----CGACTTCCCAT----- 840
Db 457 ArgAspProGlyProProGlyAlaAlaHisGlyProAlaGlyProGlyAlaHisGlyPro 476
QY 839 ACAGATTCGCGAGCGCTTTCAGAGCCCTTGAACTTCCAGAGAGGCTGTGGGCGACACAC 780
Db 477 AlaGlyProGlySerGlyAlaHisGlyProAlaGly-----ProGlySerGlyAlaHisGlyProAla 495
QY 779 AGTACCAGTGAACG---GGTGCAGCTTCTCTGTATGACGAGAGGTGGCCACACCCGAG 723
Db 496 G1yProGlySerGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyProPro 515
QY 722 ACTTGGGGTCACTCCAGACATGTGATAACCTTGGTGGCAGAAAGTCAAGCCCGCCCTCT 663
Db 516 -----GlyAlaProGly-----ProAlaGlyPro 523
QY 662 CGCCTGTCTGATTTGGCCACCCACAGTCGCTGTATTATGATGAGAGAGAAGAGCAG 603
Db 524 ProGlySerArgAspProGlyProProGly-----AlaPro 535
QY 602 GGTCCGAGAGGC-----AGATTTGGGGTCCATCCGGGCGCCCTGAGCAGCTGGG 555
Db 536 G1yProAlaGlyProProGlySerArgAspProGlyProProGlyAlaProGly----- 553
QY 554 TCTTGATTTCCAGGGGTTTCTATAGGGGACACATGAAAGCCCTTCTGGCGCGCTCCGAG 495
Db 554 -----ProAla-----GlyProProGlySer-----ArgAsp 562
QY 494 AGTGAAGAGAGCTGTCTGCTGAGCTGGAAGAGAGAGAGCCACTCCTGCTGGAAGTCTGT 435
Db 563 ProGlyPro-----ProGlyAlaProGlyProAlaGlyProProGlySerArg 578
QY 434 AGGAGTGATGCCGAGACACCCAGCGGCTTCCGCTCCAGACGCTCTCTCCGAG 375
Db 579 Asp-----ProGlyProProGly-----AlaProGlyProAlaGlyProProGly 593
QY 374 AGTGAAGACCATGTTGGGGGCTGGCTGGAATATCATCATGCTTGTCCAGACAGAGA 315
Db 594 SerArgAspProGly-----ProProGlyAlaProGlyProAlaGlyProProGly 610
QY 314 GCACAGAGAGCT-----CTTTCGAGCTTCTTGGGAATCTCAAGATGAGAGA 267
Db 611 SerArgAspProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArgAsp 630
QY 266 GGGAGTTCTCTCGGCTGC-----CATATGGCATTTCCAGTAGAGAGC----- 222
Db 631 ProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyPro 650
QY 221 ---GGTGGAGTGGGGCCAGACTCATCCGCTTCTGACACAAATGGAAGTCGGGGGG 165
Db 651 ProGlyAlaProGlyProAlaGlyProProGlySerArgAspProGlyProProGlyAla 670
QY 164 CTTGTGTGACAATGAGCGCCGAGT---ACTTGGCGGCTGCGGTGATGATGCTCGGAGCC 108
Db 671 Pro-----GlyProAlaGlyProProGlySerArgAspProGlyProProGly 686
QY 107 CGTCCACGAGTGTCTTGACACTGGAAGGGGGCGCGGCTCATCTGTGC----- 56

```

```

Db 687 AlaProGlyProAlaGlyProProGlySerArgAspProGlyPro-ProGlyAlaHisGly 706
QY 55 -----GGCTGGCTCCGCTCCGCTCCGCTGGGGTCCCGGTCGTCGACATGCT 8
Db 706 yProAlaGlyProGlySerGlyAlaHisGlyProAlaGlyProGlySerGlyAlaHisGly 724

RESULT 13
US-08-642-255-62
: Sequence 62, Application US/08642255
: Patent No. 5773249
: GENERAL INFORMATION:
: APPLICANT: CARPELLO, Joseph
: TITLE OF INVENTION: High Molecular Weight Collagen-Like
: TITLE OF INVENTION: Protein Polymers
: NUMBER OF SEQUENCES: 135
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,255
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ROWLAND, Bertram I.
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A55556-3/BIR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299 FHT UR
: INFORMATION FOR SEQ ID NO: 62:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1064 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-642-255-62

Alignment Scores:
Pred. No.: 2,57e-11 Length: 1064
Score: 253.00 Matches: 228
Percent Similarity: 27.64% Conservative: 18
Best Local Similarity: 25.62% Mismatches: 330
Query Match: 5.25% Indels: 315
DB: 1 Gaps: 48

US-09-976-674-4 (1-2617) x US-08-642-255-62 (1-1064)
QY 2516 GGTAGATCTGAGCT---GGTAAGCTTCCCTCGCATCATGTTGGAGACGAGAGT 2460
Db 103 G1yProProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyPro 122
QY 2459 TTGTGTGAAAAAGTGCACGTTTTCGTCACGAAAGCCGTGAGAGATACGACGCGTTGG 2400
Db 123 ProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGly 142
QY 2399 GCTCATTTGGCAGCGCTTCTCCA-----CGTGCA 2373
Db 143 AlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaPro 162
QY 2372 GGGCCAGGAACCGCCTCATATAGCCGTGTGTTGTTCTCAGGAGACGTCCATGTAGCGCT 2313

```

```

Db 163 GlyProAlaGlyProProGlySerArgLy-----AsProGlyProProGlyAla--- 179
Qy 2312 CAGGTACCCCTGTGTGAGGCCATCCAGAGGTGACCGGGGGACCCCGCATGGCCACT 2253
Db 180 -----ProGlyProAlaGlyProProGlySerArgLyAsPro 192
Qy 2252 TGAACACCTGGGGCTTGTGGATTAGCC-----CCATGAGCGAGAGAGGCCCGCGTAGG 2199
Db 193 GlyProProGlyAlaProGlyProAlaGlyProProGlySerArgLyAsPro----- 210
Qy 2198 ACCAGCCATGGATGGCAATCGGCTCAGGTGCATGAAAGCCATCTTCGGCCAGCAACT 2139
Db 211 -----GlyProProGlyAlaProGlyPro 218
Qy 2138 GCAGGCCCTCA-----CCGTGCTCTGCATCT 2112
Db 219 AlaGlyProProGlySerArgLyAsProGlyProProGlyAlaProGlyProAlaGly 238
Qy 2111 CCACCTGGCCCATTTGGTTTTCAGGGCCCTTCGAAACCGAAGCCCGCTGCAGAGAGC 2052
Db 239 ProProGlySerArgLyAsProGlyProProGlyAlaPro-----Gly 253
Qy 2051 CCTGCGCCATATCAACAACACAGCGGTAGCCAGGGAGGCCAGTGTGAGCCGCAAGT 1992
Db 254 ProAlaGlyProProGlySerArgLyAsProGlyPro----- 267
Qy 1991 ACTTGATCCCTTTGAGAGATTATTCACCAAGCTGACCTGGGGCCCTCATATACAAAGA 1932
Db 268 -----GlyAlaHisGlyProAlaGlyProGlyAlaHisGlyProAla 282
Qy 1931 GGAGCGTGGGTGCTTCTTCCCTGCTGCAGAGCGCTGGGGCTTGATGATCATGCCGTGA 1872
Db 283 GlyProGlySerGlyAlaHisGlyProAlaGlyProGlyAlaHis----- 296
Qy 1871 GCCGCATCCGAGGCGCTGTGAATGTGGAAGATCTGGAAGAAATATCCGGGGGCG 1812
Db 297 -----HisGlyProAlaGly 301
Qy 1811 AGTGGCTGCTCATTCATGCTAGCCCGAAGCGGGGCTGCTTGCGAGGGGTGCTGCT 1752
Db 302 -----ProGlySerGlyAlaProGlyProAlaGly----- 310
Qy 1751 CGGGGCCCTCAGCTTGTAGAAGTGCACGAGCGGGCGGTGCTACGCTGCTAGTGCC 1692
Db 310 ----- 310
Qy 1691 TGACGACATGTCGAATGTTGCTCATGAGCAGCTATGGAGAGCGGGCGTGTGA 1632
Db 311 -----ProProGlySerArgLyAsProGly----- 319
Qy 1631 GCGGTAGCATTCGCGCGCGCTCATAGCTGACCAAGTGAAGTGTGCTCCAGCGCG 1572
Db 320 -----ProProGlyAlaHisGlyProAlaGlyProProGlySer 332
Qy 1571 -----TGTCTGTGGCCCTGGAATGACACAGAGCTGTGCTTCCTCATGA 1527
Db 333 ArgGlyAsProProGlyProProGlyAlaProGly-----ProAlaGlyProPro----- 348
Qy 1526 CCCAGATCTGGAGCGCTGCGTCCGAACATCCCATTCACCGCTGCTCAGAGCAATCT 1467
Db 349 -----GlySerArgGlyAsProGlyProPro-----Gly 358
Qy 1466 CTTCCTTATATGGGCACTTAAATTCATCTCCCGGGGCTGAAGGGCTCATCAATGCT 1407
Db 359 AlaPro-----GlyProAlaGly----- 364
Qy 1406 AGCCCTGGGATTTTAAAGCGCGTGCATTTGTACAATGGCAGAGCGGCTTGCAAT 1347
Db 365 ProProGly----- 367
Qy 1346 CATTTGGCGGAGAAAGCAGAGCTGCTCCTCTCTGATTGGGGAGGATAGAGA 1287

```

```

Db 368 -----SerArgLyAsPro----- 372
Qy 1286 TGTGATGAACATGATCCAGAGCGTTGGTACCTCTCTCTATACACCAATAGCGTGCATAT 1227
Db 373 -----GlyProProGlyAlaProGlyProAlaGlyPro 383
Qy 1226 TCTGGGAGACACTCTGACAGAGCTAGCCGCTCTCTCTCATCTCTGCTCGGGATGA 1167
Db 384 ProGlySerArgLyAsProGlyProProGlyAlaPro-----GlyPro 398
Qy 1166 ACAGG---CCGGGGGAGAGAGACAGAGCTGAGCCACTGCTGGGGCGGTCCAGAGACA 1110
Db 399 AlaGlyProProGlySerArgLyAsProGlyProProGlyAlaProGlyProAla-Gl 418
Qy 1109 TGGCCAGCGCTATTTGCCATCCGGGTCCACCGGCG---CCTGGCAATGACT-CCACC 1054
Db 418 yProProGlySerArgLyAsProGlyProProGlyAlaProGlyProAlaGlyProPr 438
Qy 1053 TTGC-----GGAACAGCGAGCTGAAGGGCTGCACACACTCTCTCT-----CCTGGGTC 1006
Db 438 OGlySerArgGlyAsProProGlyProProGlyAlaProGlyProAlaGlyProProGlySe 458
Qy 1005 GAGACGATCTTGCCCTGCTGCTCAGTCTGAACCTCAGCCAGTTTCAGAGCAATCTTGGA 946
Db 458 ArgGlyAsP-----ProGlyProProGlyAlaProGlyProAlaGly----- 472
Qy 945 TTCTTGCTGCTGCTCTGGGGTACCGATACAGTCCGCTCTCTCTTTCTTAGCGCAGA 886
Db 473 -----ProProGlySer----- 476
Qy 885 GAGGGAGCTGAATACCTCCACCTCGGACATCATGACTCTCTCATACAGATTGCGAGC 826
Db 477 -ArgGlyAsProProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArgLy 496
Qy 825 GTTTGAGCGCTCTGAACCTTCCAGAGGCTGTGGGCAACCAAGTACCCAGTACCAGTGAAG 766
Db 496 yAsProGlyProProGlyAlaPro----- 504
Qy 765 CGGTGGAATCTTCTCTGATGACGAAGTGGCCACACCCCGAATTTGGGTGATCCAGC 706
Db 505 -GlyProAlaGlyPro-----ProGlySerArgLyAsProGly 517
Qy 705 ACATGTGATTAACCTTGTGGCAGAGTCAAGCCGCTCTCTCTCTCTCTGATGTTGT 646
Db 517 yPro-----ProGlyAlaHisGlyProAlaGlyProGlyAlaHis----- 531
Qy 645 GCCACCAAGTCTGCTGTTATGATGAAGAGAGAGAGGAGTGGCGAGCAGATT 586
Db 532 -GlyProAlaGly-----ProGlySerArgLyAsProGlyProAl 543
Qy 585 TTGGGCTCATTCGCGGGCTGAGC-----ACTGGCTTGTATTTCCAGCGGT 538
Db 543 a---GlyProGlySerGlyAlaHisGlyProAlaGlyProGlySerGlyAlaProGlyProAla-- 561
Qy 537 TTCATAGGGAGACCATGAAGCGCTTCTGGCGCGGTGGCGACATGGAAGAGGTGTTG 478
Db 562 -----GlyProProGlySerArgLyAsProGlyProProGlyAlaProGlyProAl 579
Qy 477 CTGGCCTGGAAGAGAGAGAGGCCACTCTGCTGGAAGTCTAGAGGTGATGCCGAAG 418
Db 579 aGlyProProGlySerArgLyAsPro-----Gl 589
Qy 417 ACCCGGAGCGTTTCCGCTCTCAGCAGCTCTCTCCCGAGATAA----- 369
Db 589 yProProGly-----AlaProGlyProAlaGlyProProGlySerArgLyAsProGly 607
Qy 368 -CCCATGGTGGGCGCTGCGCTGGAATGATCAGACACTGCTTCCAGAGCAGGA----- 315
Db 607 yProProGlyAlaProGlyProAlaGlyProProGlySerArgLyAsProProGlyProPr 627
Qy 314 -GCAGCAGAGCCTTTTCCGAGCTTCTTGGAAATCTCAGAGTAGAGAGGAGTTCTCT 256
Db 627 oGlyAlaProGlyProAlaGlyProProGlySerArgLyAsProGlyProProGlyAla 647

```



```

Db 1207 SerAlaSerGlyProAlaLeuLeuThrSerValThrProLeu----- 1221
QY 1407 ACATGTGAGTACGCTTCAGCCCGGAGAAATGATTTA-----AGTCC 1454
Db 1222 ---AlaProValValProAlaAlaProGlyProProSerLeuGlnProSerGlyAlaSer 1240
QY 1455 CCATTAAGAGAGATTGCTCTGACACGCGGTGAATGGAGTTTGGCGAGGACGGCT 1514
Db 1241 ProSerAlaSerAlaLeu-----ThrLeuGlyLeuAlaThrAla 1253
QY 1515 CCAAGATCTGGGTCAATGAGGAGACCAAGCTGTGTACTTCCAGGACACAGACAGCC 1574
Db 1254 ProSerLeuSerSerSerGlnThrProGlyHisProLeuLeuAlaProThrSerSer 1273
QY 1575 -----CGCTGAGACCAACCTCTACGTGTGACGTATGAGGCGCGCGCG 1619
Db 1274 HisValProGlyLeuAsnSerThrValAlaProAlaCysSerProValLeuValProAla 1293
QY 1620 AGATGTAACGCTTCACACGCGCGCTTCCTCCATAGCTGCTCCATGAGCAGAACTTCG 1679
Db 1294 SerAlaLeuAlaSerProPherProSerAlaProAsnProAlaPro---AlaGlnAlaSer 1312
QY 1680 ACATGTCGTACGACCTACAGACAGCTGAGCAGCGCGCTGCGGTGACACTTACAAAC 1739
Db 1313 LeuLeuAlaProAlaSerSerAlaSerGlnAlaLeu-----AlaThrProLeuAlaPro 1330
QY 1740 TGAGCGGCGCGGACGACGCGCTTCGACACAGACGCGCGCTTCGGCTGACATGATG 1799
Db 1331 MetaAlaAlaProGlnThrAlaAlaLeuAlaProSerProAlaProProAlaProLeu 1350
QY 1800 -----AGCAGACAGCTGCGCGCGATATGATTCCTCCAGATCTTCAT 1847
Db 1351 ProValLeuAlaProSerProGlyAlaAlaProValLeuAlaSerSerGlnThrProVal 1370
QY 1848 -----TCCACACGCGCTGCGATGTCGCGCTTACGGCATGATCTACA 1889
Db 1371 ProValMetaAlaProSerSerThrProGlyThrSerLeuAlaSerAlaSerProValPro 1390
QY 1890 AGCCCGACGCTTCGACGACGAGGAAAGAACCCCA-----CCG 1928
Db 1391 AlaProThrProValLeuAlaProSerSerThrGlnThrMetLeuProAlaProValPro 1410
QY 1929 TCCCTTTGATATGAGAGCGCCCGACAGTGTGATGATCTTCCTCAAGAGCATCA 1988
Db 1411 SerProLeuProSerProAlaSerThrGlnThrLeuAlaAlaProAlaLeuAlaPro 1430
QY 1989 AGTACTTGGCGCTCAACACACTGCGCTCCCTGGGCTACGCGGTGATGATGACGCA 2048
Db 1431 ThrLeuGlyGlySerSer-----ProSerGlnThr-----LeuSerLeuGlyThr 1445
QY 2049 GGGGCTCTGTTCACGAGGAGGCTTGCTGCGAGGCGCCCTCAAAAACCAATGGCCAG 2108
Db 1446 GlysAsnPro-----GlnGlyProPheProThr----- 1454
QY 2109 TGGAGATCGAGACGAGGTGAGGCGCTTCAGTGTGCGGAGAGATGAGCTTCATCG 2168
Db 1455 -----GlnThrLeuSerLeuThrProAlaSerSerLeuValPro 1467
QY 2169 ACCTGAGCGAGTTCATCATGAGCT-----GGTCTACGCGGGCTTC 2213
Db 1468 ThrProAlaGlnThrLeuSerLeuAlaProGlyProProLeuGlyProThrGlnThrLeu 1487
QY 2214 TCTGGCTCATGGGCTATTCACAAAGCCCAAGGTGTCAAGGTGCGCATCGCGGGTGGCC 2273
Db 1488 -----SerLeuAlaProAla-----ProProLeuAlaPro 1497
QY 2274 CGGTACCGCTGTGATGCGCTACGACACAGGTTACATG----- 2312
Db 1498 AlaSerProValGly---ProAlaProAlaHisThrLeuThrLeuAlaProAlaSerSer 1516
QY 2313 AGCGCTACATG-----ACGTCCCTGAGAACACACAGCAGCGCTATGAGCGGGTT 2363
Db 1517 SerAlaSerLeuLeuAlaProAlaSerValGlnThrLeuThrLeuSerProAlaProVal 1536

```

```

QY 2364 CCG-----TGGCCCTGACGCTGGAGAACTGCCAATGACCCCAACCGCTTGCTATCC 2417
Db 1537 ProThrLeuGlyProAlaAlaAlaGlnThrLeuAlaLeuAlaPro---AlaSerThrGln 1555
QY 2418 TCCAGCGGCTTCCTGAGCAAAACGTGACCTTTTCC----- 2453
Db 1556 SerProAlaSerGlnAlaSerSerLeuValSerAlaSerGlyAlaAlaProLeuPro 1575
QY 2454 ---ACACAACTTCCTGCTCTCCCACT-----GATCCGAGCAGGAAACCTTACC 2501
Db 1576 ValThrMetaValSer-ArgLeuProValSerLysAspGluProAspThrLeuThrLeuAr 1595
QY 2502 AGCT-----CCAGATCTACCCCAACGAGACACAGATTCGCTGCCGCGAGCGGCG 2555
Db 1595 gSerGlyProProSerProProSerThrAlaThrSerPheGlyProArgProArgAr 1615
QY 2556 A 2556
Db 1615 g 1615

RESULT 15
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCY
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

Alignment Scores:
Pred. No.: 4,15e-11 Length: 657
Score: 249.00 Matches: 152
Percent Similarity: 35.77% Conservative: 112
Best Local Similarity: 20.60% Mismatches: 246
Query Match: 5.19% Indels: 228
DB: Gaps: 34

US-09-976-674-4 (1-2617) x US-09-355-166-1 (1-657)
QY 545 GAATCAAGACCCAGTCTGACAGGCCC-----CGATGAGACCC 583
Db 50 GlnThrLysThrGlnGlyLysValAlaProThrPThrHisGlyGluLysArgSerThrAspPro 69
QY 584 AAATCTCCCTGCGCGACCTGCTCTTCCTTCATCAATAC----- 628
Db 70 ArgTrpSerPro---AspGlyArgThrLeuAlaPheIleSerAspArgGluLysPala 88
QY 629 AGCAGCTGTGGGTGGCCACATCGAGACAGCGAGGAGCGGCGCTGACCTTGGCAC 688
Db 89 AlaGlnLeuTyrlIleMetSerThrGlnGlyGluAlaArgLysLeuThrAspIlePro 108
QY 689 CAAGTTTATCAATGTCCTGATGACCCCAAGTCTGGGGGTGGCCACCTTCCTCATTA 748
Db 109 TyrGlyValSerLysProLeu----- 115
QY 749 CAGGAAGATTGACCGCTTCACTGGTACTGTGTCGCCACACGCTCCTGGAGAGT 808
Db 116 -----TrpSerPro-----AspGly 120
QY 809 TCAGAGGCGCTCAAGACGCTGGGAATC---CTGTATGAGGAATCGATGATCGAGGTG 865
Db 121 GluSerIleLeuValThrIleSerLeuGlyGluSerIleAspAspArg----- 138

```



```

QY 866 GAGGTATTCACTGCTCCCTCTCTGCGTAGAAGAAAGACGAGCTGATCGGTAC 925
    |||::: |||
Db 139 -----GluysThrGlnGlnAspSerTyrGlu--- 147
QY 926 CCCAGGACGACGACGACAGATGCCAAGATTGCCCTGAAGCTGGGTGATCCAGACTGAC 985
    |||::: |||
Db 148 -----ProValGlnValGlnGlnLeuSerTyrLysArgAsp 159
QY 986 AGCCAGAGGC-----AAGATCGTCTCGACC 1009
    |||::: |||
Db 160 GYLysGlnLeuThrArgGlnAlaTyrAlaGlnLeuValLeuValSerValLysSerGly 179
QY 1010 CAGAGACAGAGAGCTGGTGGACAGCCCTTCACTGCTGTTCCCGAAGTGGAGTACATCGCC 1069
    |||::: |||
Db 180 GluMetLysGlnLeuThrSer-----HisLysAlaAspHis---Gly 192
QY 1070 AGGGCCGGGTGGACCCCGGATGGCAATAGCCCTGGGCCATGTTCCCTGGACCGGCCACG 1129
    |||::: |||
Db 193 AspProAlaPheSerProAspGlyLys-----TyrLeuValPhe----- 205
QY 1130 CAGTGGCTCAGACTCGTCTCTGCTCCCGCCGCGCTGTTCACTCCGACGACAGAGATGAG 1189
    |||::: |||
Db 206 -----SerAlaAsnLeuThr 210
QY 1190 GAGCAGCGGCTAGCCTCTGCTCCAGAGCTGTCCGAGAAATGCCAGCCGATGTGGTGTAC 1249
    |||::: |||
Db 211 GluThrAspAspAlaSer----- 216
QY 1250 GAGGAGGTGCACCAACGCTGGATGCATATGTTGATGACATCTTCTTACCTCCCTCCCAATCA 1309
    |||::: |||
Db 217 -----LysProHisAspValTyr----- 222
QY 1310 GAGGAGAGAGACAGCTCTGCTTCTCCGCGCAATGAAATGACAAAGCCGCTTGTGCCAT 1369
    |||::: |||
Db 223 -----IleMetSerLeuGlnSerGlyAspLeuLys----- 232
QY 1370 TTGTACAAAGTCAACCCGCTTTTAAATCCAGAGGCTACAGATGGAGTACAGCCCTTCAGC 1429
    |||::: |||
Db 233 -----GlnValThrProHisArgGlySerPheGlySerSerPheSer 247
QY 1430 CCC-----GGGGAAGATGAATTTTAAGTGTCCCATTAAG 1462
    |||::: |||
Db 248 ProAspGlnArgTyrLeuAlaLeuAlaLeuGlnLysGlnLysGlnLys----- 263
QY 1463 GAAGAGATTGCTGTGACACGACGCTGAATGG----- 1492
    |||::: |||
Db 264 -----AsnAlaThrLeuSerLysAlaTyrLeuTyrAspIleGlnGlnLysArgLeuThr 281
QY 1493 -----GAGTTTTGGCGAGGCAC----- 1510
    |||::: |||
Db 282 CysLeuThrGlnMetLeuAspValHisLeuAlaAspAlaLeuIleGlyAspSerLeuIle 301
QY 1511 ---GGCTCCAG-----ATCTGGGTCAATAGAGAGACCAAGCGGTGTACTTCAG 1558
    |||::: |||
Db 302 GlyGlnAlaGlnLeuAlaArgProIleTyr---ThrLysAspSerGlnGlyPheTyrValIle 320
QY 1559 GGCACCAAGACACGCGCTGGAGACACCACTAGCTAGCTGAGTATGAGCGCGCGC 1618
    |||::: |||
Db 321 GALTThrAspGlnIle-----SerThrGlnIleTyrTyrIleSerIleGlnGlnLeuVal 338
QY 1619 GAGATGTTAGCTCAGACACGCGCGCTGTCTCCCATAGCTGTCCATGAGCCAGAACTTC 1678
    |||::: |||
Db 339 TyrProIleArgProIleLysGln---TyrIleAsnSerPheSerLeuSerProAspGln 357
QY 1679 GACATGTTGCTGACGACATACACACGCTGAGACACCGCCCTGCGTGCAGCTTACAG 1738
    |||::: |||
Db 358 GlnHisPheThrLeuAlaSerValThrLysProAspArgProSer-----GluLeuTyrSer 375
QY 1739 CTGAGGCGGCGCCGACAGACGCGCTGTGACAAAGACGCGCTTGGGTAGCATGATG 1798
    |||::: |||
Db 376 Ile-----ProLeuGlnGlnGlnGlnGlnLysGlnLysGlnLysGlnLysGlnLys 389
QY 1799 GAGGACGACGAGCTGCGCCCGGATGATGTTCTCCAGAGATCTTCATTTCCACACGCGC 1858

```

```

Db 390 AspLysPheValArgGlnHisThrIleSerIleProGlnGlnIleGlnTyrAlaThrGlu 409
    |||::: |||
QY 1859 TCGGATGTGGCGCTCTAGCCGATGATCTACAGCCCGCCGCTTGGACGACGAGAAAG 1918
    |||::: |||
Db 410 AspGlnValMetValAsnGlyTyrPheMetArgProAlaGlnMetGlnGlnGlnThr 429
QY 1919 CAGCCGACGCTCTCTTGTATATGAGAGCCCGCAGGTGACGCTGGAAATTAACCTTC 1978
    |||::: |||
Db 430 TyrProLeuIleLeuAsnIleHisGlyGlyProHisMetMetTyrGlnHisThrPhe 449
QY 1979 AAGGACATCAAGTACTTCCGCTCACACACTGCGCTCCCTGGGCTAGCGCTGGTGTG 2038
    |||::: |||
Db 450 His-----GluPheGlnValLeuAlaLysGlnTyrAlaValValTyr 464
QY 2039 ATTACAGCGAGGCGCTCTCTAGCGAGGCGCTTGGTTCGAAAGGCGCCGAAACCAA 2098
    |||::: |||
Db 465 IleAsnProArgGlnSerHisGlyTyrGlnGlnPheValAsnAlaValArgGlyAsp 484
QY 2099 ATGGGCGAGGTGGAGATCGAGACACAGTGGAGGCGCTGCGAGTTCTGGCGGAAATAT 2158
    |||::: |||
Db 485 TyrGlnGlnLysAspTyrArgAspValMetGlnAlaValAspGlnAlaIleLysArgAsp 504
QY 2159 GCGTTCATGCACTGACGAGCTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2218
    |||::: |||
Db 505 ProHisIleAspProLysArgLeuGlnValThrGlnGlnLysGlnLysGlnLysGlnLys 524
QY 2219 -----CTCATGGGCGCTAATCCACACAGCCCGAGGTGTTCAAGTGGCGCAATGCGCGGTGCC 2272
    |||::: |||
Db 525 AsnTyrPheValGlnGlnThrAsnArg-----PheLysAlaAlaValAlaThrGlnArg 541
QY 2273 CCGGTACGCTGTGATGAGCTAC-----GACACAGGCTAC-----ACTGAG 2314
    |||::: |||
Db 542 SerIleAsnThrIleSerPheHisGlyValSerIleGlyTyrPhePheThrAsp 561
QY 2315 CGCTACATG-----GAGTCCCTGAGACACACGACGCTATGAGCGGCTTCCGTG 2368
    |||::: |||
Db 562 TyrGlnLeuLysHisAspMetPheGlnAspThrGlnLysLeuTyrAspArgSerProLeu 581
QY 2369 -----GCCGTGACGCGGAGAGAGCTGCCCAATGAGCCCAACGCTGTTATCTTCAC 2422
    |||::: |||
Db 582 LysTyrAlaAlaAsnValGln-----ThrProLeuLeuIleLeuHis 595
QY 2423 GCGTCTGAGCAGAAACGTCGACTTTTCCACAAACTTCTGCTGCTGCTGCTGCTGCTGCTG 2482
    |||::: |||
Db 596 GlnGlnArgAspArgCysProIleGlnGlnAlaGlnGlnLeuPheThrLeuValLys 615
QY 2483 CGAGCAGGGAACCTTACAGCTTCAGATCTACCCCAAGCAGACAGACATAT 2536
    |||::: |||
Db 616 LysMetGlnLysGlnThrLysLeuValArgPheProAsnAlaSerHisAsnLeu 633

```

Search completed: December 12, 2002, 11:54:35
 Job time : 63.5 secs

Warren, B.A., Lu, D.A., Lee, E.A., Tribouley, C.M., Arvizu, C.,
Deleane, A.M., Yao, M.G., Khan, F.A. and Sanjanwalla, M.M.
Protein modification and maintenance molecules
Patent: WO 0246383-A 30.13-JUN-2002;
Incyte Genomics, Inc. (US)

FEATURES
source Location/Qualifiers

1..3716
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7972712CB1"
BASE COUNT 768 a 1174 c 1075 g 699 t
ORIGIN

Query Match 99.1%; Score 2592.4; DB 6; Length 3716;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 ATGGCCACACCGGGAGCCCAACGCGCGACGAGCGAGCGCGCCACAGATGACCG 70
DB 330 ATGGCCACACCGGGAGCCCAACGCGCGACGAGCGCGCGCCACAGATGACCG 389
QY 71 GCGCGCGCTTCAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 130
DB 390 GCGCGCGCTTCAGGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 449
QY 131 AGCGCAAGTACTGCGGCTCATTTGTCACAGAGCGCGCCACGACTTCAGTTTGTCAG 190
DB 450 AGCGCAAGTACTGCGGCTCATTTGTCACAGAGCGCGCCACGACTTCAGTTTGTCAG 509
QY 191 AAGACGAGTATGCTGGGCGCCACTCCACCGCTTACTACTGAGGAGTATGATGATG 250
DB 510 AAGACGAGTATGCTGGGCGCCACTCCACCGCTTACTACTGAGGAGTATGATGATG 569
QY 251 AGCGGAGAGTATGCTGGGCGCCACTCCACCGCTTACTACTGAGGAGTATGATGATG 310
DB 570 AGCGGAGAGTATGCTGGGCGCCACTCCACCGCTTACTACTGAGGAGTATGATGATG 629
QY 311 CTGCTCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 370
DB 630 CTGCTCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 689
QY 371 TACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 430
DB 690 TACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 749
QY 431 TCTAGACTTTCACAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 490
DB 750 TCTAGACTTTCACAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 809
QY 491 CACTGTGCGAGCGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 550
DB 810 CACTGTGCGAGCGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 869
QY 551 AAGACCGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 610
DB 870 AAGACCGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 929
QY 611 TTCTCTTATCATATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 670
DB 930 TTCTCTTATCATATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 989
QY 671 CGGCTGAGCTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 730
DB 990 CGGCTGAGCTTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1049
QY 731 GTGGCCACCTTCTGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 790
DB 1050 GTGGCCACCTTCTGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1109
QY 791 ACAGCGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 850
DB 1110 ACAGCGCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1169

QY 851 GATGAGTCCGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 910
DB 1170 GATGAGTCCGAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1229
QY 911 GACTCGTATGAGTATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 970
DB 1230 GACTCGTATGAGTATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1289
QY 971 GAGTTCAGACTGACAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1030
DB 1290 GAGTTCAGACTGACAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1349
QY 1031 CCTTCAGCTGCTGTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1090
DB 1350 CCTTCAGCTGCTGTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1409
QY 1091 GGCAGATACGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1150
DB 1410 GGCAGATACGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1469
QY 1151 CTGCCCCCGGCGCTGTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1210
DB 1470 CTGCCCCCGGCGCTGTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1259
QY 1211 AGAGCTGTCCAGAGATGTCAGCCGATGAGTATGAGGAGGAGGAGGAGGAGGAGG 1270
DB 1530 AGAGCTGTCCAGAGATGTCAGCCGATGAGTATGAGGAGGAGGAGGAGGAGGAGGAGG 1589
QY 1271 ATCAATTTATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1330
DB 1590 ATCAATTTATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1649
QY 1331 TTTCTCGGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1390
DB 1650 TTTCTCGGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1709
QY 1391 TTAAGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1450
DB 1710 TTAAGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1769
QY 1451 TGGCCCCCTTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1510
DB 1770 TGGCCCCCTTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1829
QY 1511 GGCCTCAAGATCTGGGCTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1570
DB 1830 GGCCTCAAGATCTGGGCTCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1889
QY 1571 ACGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1630
DB 1890 ACGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1949
QY 1631 CTCACGAGGCGCGGCTTCTCCATAGCTGCTCAGATGAGGAGGAGGAGGAGGAGGAGG 1690
DB 1950 CTCACGAGGCGCGGCTTCTCCATAGCTGCTCAGATGAGGAGGAGGAGGAGGAGGAGG 2009
QY 1691 AGCCACTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1750
DB 2010 AGCCACTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2069
QY 1751 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1810
DB 2070 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2129
QY 1811 TGGCCCCCGGAGTATGTTCTTCCAGAGATGTTCTTCCAGAGATGTTCTTCCAGAGAT 1870
DB 2130 TGGCCCCCGGAGTATGTTCTTCCAGAGATGTTCTTCCAGAGATGTTCTTCCAGAGAT 2189
QY 1871 CTCTACGAGATGATCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1930
DB 2190 CTCTACGAGATGATCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2249
QY 1931 CTCTTTGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1990

|||||
Db 2250 CTCCTTATATGAGGAGCCGCCAGGTGCTGTAATACTCTCAAGCATCAAG 2309
Oy 1991 TACTTGGGGCTCAACACACTGGCCCTCCCTGGCTACGCCCGGTGATTGAGGCGAG 2050
Db 2310 TACTTGGGGCTCAACACACTGGCCCTCCCTGGCTACGCCCGGTGATTGAGGCGAG 2369
Oy 2051 GGCTCTGTGACGAGGAGGCTTGGTGAAGGGGCGCTGAAGAAACCAATGGGCGAGTG 2110
Db 2370 GGCTCTGTGACGAGGAGGCTTGGTGAAGGGGCGCTGAAGAAACCAATGGGCGAGTG 2429
Oy 2111 GAGATGAGAGACACAGGTGAGAGGCGCTGAGTTCGTGGCGGAGATATGAGCTTCATCGAC 2170
Db 2430 GAGATGAGAGACACAGGTGAGAGGCGCTGAGTTCGTGGCGGAGATATGAGCTTCATCGAC 2489
Oy 2171 CTGAGCGGAGTGGCTCATCATGCTGCTCTACGAGGGGCTTCTCTGCTATGAGGCGTA 2230
Db 2490 CTGAGCGGAGTGGCTCATCATGCTGCTCTACGAGGGGCTTCTCTGCTATGAGGCGTA 2549
Oy 2231 ATCCAGACAGCCCGAGTGTTCAGGTGAGGCGCTGAGGCGGCGCTGACCGTGGATG 2290
Db 2550 ATCCAGACAGCCCGAGTGTTCAGGTGAGGCGCTGAGGCGGCGCTGACCGTGGATG 2609
Oy 2291 GCCTACGACACAGGGTACACTGAGCGGCTACATGAGACGTCCTGAGAACACACAGCAGCGC 2350
Db 2610 GCCTACGACACAGGGTACACTGAGCGGCTACATGAGACGTCCTGAGAACACACAGCAGCGC 2669
Oy 2351 TATGAGCGGGGTTCCGTGGCGCTGACGTGAGAGAGTCCCAATGAGCCCAACCGCTTG 2410
Db 2670 TATGAGCGGGGTTCCGTGGCGCTGACGTGAGAGAGTCCCAATGAGCCCAACCGCTTG 2729
Oy 2411 CTATCTCTCCACGCGCTTCTGAGAGAAAGTGACATTTTCCACACAACTTCTCTGTC 2470
Db 2730 CTATCTCTCCACGCGCTTCTGAGAGAAAGTGACATTTTCCACACAACTTCTCTGTC 2789
Oy 2471 TCCCACTGATCCGAGGAGGAGAACTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2530
Db 2790 TCCCACTGATCCGAGGAGGAGAACTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2849
Oy 2531 AGTATTTGGCTGCGCCGAGTGGGGGAGCAGCATATGAAGTCACTGCTGCTATGACG 2590
Db 2850 AGTATTTGGCTGCGCCGAGTGGGGGAGCAGCATATGAAGTCACTGCTGCTATGACG 2909
Oy 2591 GAATACCTGTGAGC 2604
Db 2910 GAATACCTGTGAGC 2923

RESULT 2
AF452102 4295 bp mRNA linear PRI 08-JUL-2002
LOCUS Homo sapiens dipeptidyl peptidase-like protein 9 mRNA, complete
DEFINITION
ACCESSION AF452102
VERSION AF452102.2 GI:21704252
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Olsen, C. and Magtman, N.
TITLE Identification and characterization of a novel member of the
dipeptidyl peptidase IV-related family
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4295)
AUTHORS Olsen, C. and Magtman, N.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Biotechnology, Novo Nordisk A/S, Novo Alle,
Bagsvaerd DK-2880, Denmark
REFERENCE 3 (bases 1 to 4295)
AUTHORS Olsen, C. and Magtman, N.
TITLE Direct Submission

JOURNAL Submitted (08-JUL-2002) Biotechnology, Novo Nordisk A/S, Novo Alle,
Bagsvaerd DK-2880, Denmark
REMARK Sequence update by submitter
COMMENT On Jul 8, 2002 this sequence version replaced gi:17865310.
FEATURES
source
Location/Qualifiers
1..4295
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
366..2957
/note="Dipep9; peptidase family 9b; LOC91039; HS.237617"
/codon_start=-1
/product="dipeptidyl peptidase-like protein 9"
/protein_id="AA147179.1"
/db_xref="GI:17865311"
/translation="MATTPPTDRGDAAATDPAARFOYOKHSMGLRSIIHGRKY
SGILVNKAPHDFOFVOKTDEGSHSLYLYLAMPYSGRENSLYSEIPIKVRKALIL
LSMKOMLDFHOATPHGVSREBELERRRLGVITSDYDFHESGLFLFOANSIF
HCRDGGKNGFMVSPMKLEIKTOSGSRMDPKICPADAPAFSPFINNSDLVANIEGE
ERRITPCHOGLSNVLDPKSAGAVTPIQIEPDRFTYMWCPITASMESGKTLIRL
YEYDESEVEYIHYPSALDERKTDSTRIYRTGSKNPKIALKLAEPOTDSOGKIVSTO
EKELVQPFSSLPKVEYIARAGWTRDKYAMAFLDPOQMLDVLPLPALFIPSTEN
EEOPLASARAVPRNVQPYVYEEVYINWYHDIFFPQSEGEDELCTIRANEKKG
FCHLYKTVAVLKSGYDMSEPFSGDEPFCKIIEELATSGEVEYLARHSGKIWNE
ETRLVYRQGRKIDPLRHLVYVEAAGEIVYRLTTPBFSHSCSMOYFMDVSHYSV
STPCVAVYRLSGDDPDLKOPRFAMSMBAASCPDYIPPELFIHHTSDYRLYRM
IYRPHALQPKRHEPTLVLYGQPOVOLVNNFSFGIKRLYRLNTLASLYAVAVIDGRS
CORLRFKVALIKQMGVEIEDQVEGLQFVAKEGFIDLSRYALHSGSYGFLSLMGL
IHRPOYFKAIIAGAPYVMAAYDTEGYERMDYDENNOHSEYAGSVLAHVEKLPENR
RLILHGFLENDVHFHTNPLVSQLINAGRPYQIYIPNERHRSIRCESEBHEVTL
HPIQEVY"
BASE COUNT 922 a 1346 c 1220 g 807 t
ORIGIN
Query Match 99.1%; Score 2592.4; DB 9; Length 4295;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 11 ATGGCCACCACCGGAGACCCCAACGCGCGAGCGAGCGACGCCACAGATGACCG 70
Db 366 ATGGCCACCACCGGAGACCCCAACGCGCGAGCGAGCGACGCCACAGATGACCG 425
Oy 71 GCCGCGCGCTTCCAGGTGACAGACACTGTTGGGAGCGGCGTCCGGAGCATATCCAGCG 130
Db 426 GCCGCGCGCTTCCAGGTGACAGACACTGTTGGGAGCGGCGTCCGGAGCATATCCAGCG 485
Oy 131 AGCCGCAAGTACTGCGGCTCATTTGTCAACAAGCGGCCCAAGACTTCCAGTTGTGCGAG 190
Db 486 AGCCGCAAGTACTGCGGCTCATTTGTCAACAAGCGGCCCAAGACTTCCAGTTGTGCGAG 545
Oy 191 AAGAGGATGAGTGTGGGCCCACTCCACCGGCTCTACTACCTGGAATGCCATATGCG 250
Db 546 AAGAGGATGAGTGTGGGCCCACTCCACCGGCTCTACTACCTGGAATGCCATATGCG 605
Oy 251 AGCCGAGAGACTCCCTCTACTCTGAGANTTCCCAAGAGTCCGGAAGAGGCTCTG 310
Db 606 AGCCGAGAGACTCCCTCTACTCTGAGANTTCCCAAGAGTCCGGAAGAGGCTCTG 665
Oy 311 CTGCTCTGCTCTGGAACAGATGCTGATATTTCCAGGCGACGCCACCATGGGGTGC 370
Db 666 CTGCTCTGCTCTGGAACAGATGCTGATATTTCCAGGCGACGCCACCATGGGGTGC 725
Oy 371 TACTCTCGGAGAGAGACTCTGAGGAGCGGAAACCGCTGGGGTCTTGGCATCAC 430
Db 726 TACTCTCGGAGAGAGACTCTGAGGAGCGGAAACCGCTGGGGTCTTGGCATCAC 785
Oy 431 TCTTACAGACTTCCACAGGAGAGAGTGGCTTCTCTTCCAGGCGACAGAAAGCTCTTC 490
Db 786 TCTTACAGACTTCCACAGGAGAGAGTGGCTTCTCTTCCAGGCGACAGAAAGCTCTTC 845
Oy 491 CACTGTGCGAGCGGCGGAGAAAGCGCTTCAATGATGCTCCCTATGAACCGCTGAATTC 550

Db 846 CACTGCCGCGACGCGGCAAGAACGGCTTCATGATGTCCTCCATATGAACCGCTGGAAATC 905
QY 551 AAGACCCAGTGTCTAGAGGCCCCCGGATGGACCCCAAAATGTGCCCTGGCAGACCTGCTTC 610
Db 906 AAGACCCAGTGTCTAGAGGCCCCCGGATGGACCCCAAAATGTGCCCTGGCAGACCTGCTTC 965
QY 611 TTCTCTTCATCATTAACAGCGACCTGTGGTGGCCAAACATCGAGACAGGCGAGGACGG 670
Db 966 TTCTCTTCATCATTAACAGCGACCTGTGGTGGCCAAACATCGAGACAGGCGAGGACGG 1025
QY 671 CGGCTAGACCTTCTGCCACCAAGGTTTATCCAAATGCTCTGGATGACCCCAAGTCTGGGGT 730
Db 1026 CGGCTAGACCTTCTGCCACCAAGGTTTATCCAAATGCTCTGGATGACCCCAAGTCTGGGGT 1085
QY 731 GTGGCCACCTTCTGTCATACAGGAAGATTGACGCGCTTCACTGGGATCTGTTGGTGGCCCC 790
Db 1086 GTGGCCACCTTCTGTCATACAGGAAGATTGACGCGCTTCACTGGGATCTGTTGGTGGCCCC 1145
QY 791 ACAGCCTCTGGGAAGGTTTCAGAGGGCCTCAAGAGCTGGCAATCTGTATGAGAAATC 850
Db 1146 ACAGCCTCTGGGAAGGTTTCAGAGGGCCTCAAGAGCTGGCAATCTGTATGAGAAATC 1205
QY 851 GATGAGTCGAGGGTGGAGGTCATTCACGTCCCTCTGCTGCGTGAAGAAAGAAAGAACG 910
Db 1206 GATGAGTCGAGGGTGGAGGTCATTCACGTCCCTCTGCTGCGTGAAGAAAGAAAGAACG 1265
QY 911 GACTGTATCGGTACCCACAGAGCAGCAGAGATCCCAAGATTGCTTGAACCTGGCT 970
Db 1266 GACTGTATCGGTACCCACAGAGCAGCAGAGATCCCAAGATTGCTTGAACCTGGCT 1325
QY 971 GAATTCACAGCTGACAGCCAGGCGAGAGATGCTCTGACCCAGAGAAAGAGCTGTGCAG 1030
Db 1326 GAATTCACAGCTGACAGCCAGGCGAGAGATGCTCTGACCCAGAGAAAGAGCTGTGCAG 1385
QY 1031 CCGTTCAGCTCGGTTCCGGAAGGTGGAGTACATCGCCAGGGCGGGTGGACCCGGGAT 1090
Db 1386 CCGTTCAGCTCGGTTCCGGAAGGTGGAGTACATCGCCAGGGCGGGTGGACCCGGGAT 1445
QY 1091 GGCAAATACGCTGAGGCGCATGTTCTGTGACCGGCCAGCAGTGGCTCAGCTGCTCTC 1150
Db 1446 GGCAAATACGCTGAGGCGCATGTTCTGTGACCGGCCAGCAGTGGCTCAGCTGCTCTC 1505
QY 1151 CTGCCCCCGGCGCTTTTATCCGAGCAGAGAGATGAGAGCAGCGGCTTAGCCTTGCC 1210
Db 1506 CTGCCCCCGGCGCTTTTATCCGAGCAGAGAGATGAGAGCAGCGGCTTAGCCTTGCC 1565
QY 1211 AGAGCTGTCCCGAGGATGTCCAGCCGTATGTGTGTGACGAGGAGGTCAACAAGTGTGG 1270
Db 1566 AGAGCTGTCCCGAGGATGTCCAGCCGTATGTGTGTGACGAGGAGGTCAACAAGTGTGG 1625
QY 1271 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGCAGCTTGC 1330
Db 1626 ATCAATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGCAGCTTGC 1385
QY 1331 TTCTCTCCGCGCCATGATGCAAGACGGCTTCTGCCATTTGTACAAAGTCAACCGCGTT 1390
Db 1686 TTCTCTCCGCGCCATGATGCAAGACGGCTTCTGCCATTTGTACAAAGTCAACCGCGTT 1745
QY 1391 TTAAATTCACAGGGCTACGATTTGAGTGAAGCCTTCAGCCCGGGGGAAGATGATTTAAG 1450
Db 1746 TTAAATTCACAGGGCTACGATTTGAGTGAAGCCTTCAGCCCGGGGGAAGATGATTTAAG 1805
QY 1451 TGCCCATTTAAGGAAGAGATTGCTGTGACAGCGGTGAATGGAGAGTTTGGGAGGAC 1510
Db 1806 TGCCCATTTAAGGAAGAGATTGCTGTGACAGCGGTGAATGGAGAGTTTGGGAGGAC 1865
QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGACTTCCAGGGCACCAGAGAC 1570
Db 1866 GGCCTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGACTTCCAGGGCACCAGAGAC 1925
QY 1571 ACGCCCTGGAGACCAAGCTCTAGGTGTGAGTATGAGGCGGCGGAGATCGTACGC 1630
Db 1926 ACGCCCTGGAGACCAAGCTCTAGGTGTGAGTATGAGGCGGCGGAGATCGTACGC 1985

QY 1631 CTCACACGCGCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTGCATGTTCTC 1690
Db 1986 CTCACACGCGCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTGCATGTTCTC 2045
QY 1691 AGCCACTACAGACGCTGACACGCGCCTTGCCTGACGCTTACAAAGCTGAGCGGCC 1750
Db 2046 AGCCACTACAGACGCTGACACGCGCCTTGCCTGACGCTTACAAAGCTGAGCGGCC 2105
QY 1751 GACGAGACCCCTGCAACAGACGCGCGCTTGGGGCTAGCATGAGGAGGACCGAGC 1810
Db 2106 GACGAGACCCCTGCAACAGACGCGCGCTTGGGGCTAGCATGAGGAGGACCGAGC 2165
QY 1811 TGCCCCCGGATTTATGTTCCCTCCAGAGATCTTCATTTCCACACGCGCTGAGTTCGG 1870
Db 2166 TGCCCCCGGATTTATGTTCCCTCCAGAGATCTTCATTTCCACACGCGCTGAGTTCGG 2225
QY 1871 CTCTAGGGCATGATCTACAAAGCCACGCTTGCAGCAGCAGGAAGAACACCCACGTC 1930
Db 2226 CTCTAGGGCATGATCTACAAAGCCACGCTTGCAGCAGCAGGAAGAACACCCACGTC 2285
QY 1931 CTCTTTGATATGAGGCGCCCAAGGTGACAGTGGTGAATTAACCTTCAAGGATCAAG 1990
Db 2286 CTCTTTGATATGAGGCGCCCAAGGTGACAGTGGTGAATTAACCTTCAAGGATCAAG 2345
QY 1991 TACTTGGGCTCAACACACTGGCTCCCTGGGCTACGCGCTGTGTGTGATGACGCGAG 2050
Db 2346 TACTTGGGCTCAACACACTGGCTCCCTGGGCTACGCGCTGTGTGTGATGACGCGAG 2405
QY 2051 GGCCTCTGTACAGGAGGCTTCGTTGAAAGGGCCCTGAAAAACCAATAGGCGCAGTG 2110
Db 2406 GGCCTCTGTACAGGAGGCTTCGTTGAAAGGGCCCTGAAAAACCAATAGGCGCAGTG 2465
QY 2111 GAGATGAGGACCAAGGTGAGGGGCTGCAGTTCGTGGCCGAGAAATGTGGCTTCATGAC 2170
Db 2466 GAGATGAGGACCAAGGTGAGGGGCTGCAGTTCGTGGCCGAGAAATGTGGCTTCATGAC 2525
QY 2171 CTGAGCGAGTTGCCATTCATGCTGTGCTTACGCGGGCTTCTCTGCTCATGAGGCTA 2230
Db 2526 CTGAGCGAGTTGCCATTCATGCTGTGCTTACGCGGGCTTCTCTGCTCATGAGGCTA 2585
QY 2231 ATCCACAGCCCCAGGTGTTCAAGGTGGCCATGCGCGGGTCCCGGCTACCGTCTGATG 2290
Db 2586 ATCCACAGCCCCAGGTGTTCAAGGTGGCCATGCGCGGGTCCCGGCTACCGTCTGATG 2645
QY 2291 GCCCTAGACACAGGCTACCTGAGCGCTACATGAGACGCTCCCTGGAACAACAGCAGGCG 2350
Db 2646 GCCCTAGACACAGGCTACCTGAGCGCTACATGAGACGCTCCCTGGAACAACAGCAGGCG 2705
QY 2351 TATGAGGCGGGTTCCGTGGCCCTGACAGTGAAGAGCTGCCCAATGAAGCCCAAGCTTG 2410
Db 2706 TATGAGGCGGGTTCCGTGGCCCTGACAGTGAAGAGCTGCCCAATGAAGCCCAAGCTTG 2765
QY 2411 CTTATCTCTCAGCGGCTTCTGAGAGAAAGCTGACACTTTTCCACACAAACTTCTGTC 2470
Db 2766 CTTATCTCTCAGCGGCTTCTGAGAGAAAGCTGACACTTTTCCACACAAACTTCTGTC 2825
QY 2471 TTCCAACCTGATCGAGAGGAAACCTTACAGCTCCAGATCTAACCCACGAGAGAC 2530
Db 2826 TTCCAACCTGATCGAGAGGAAACCTTACAGCTCCAGATCTAACCCACGAGAGAC 2885
QY 2531 AGTATTCGTGCCCCGAGTGGGCGAGCACTATGAAGTCAAGTTGTGTCAAG 2590
Db 2886 AGTATTCGTGCCCCGAGTGGGCGAGCACTATGAAGTCAAGTTGTGTGTCAAG 2945
QY 2591 GAATTAACCTTGAGC 2604
Db 2946 GAATTAACCTTGAGC 2959

RESULT 3
HSM805448
LOCUS HSM805448 3243 bp mRNA linear PRI 12-JUL-2002

```

DEFINITION Homo sapiens mRNA; cDNA DKFp762M2413 (from clone DKFp762M2413).
ACCESSION AL834376
VERSION AL834376.1 GI:21740044
KEYWORDS
SOURCE
ORGANISM human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 3243)
REFERENCE Ansoorge, W., Warkner, U., Mewes, H.W., Well, B. and Wiemann, S.
AUTHORS Direct Submission
TITLE Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
JOURNAL Research Center for S. Wiemann, Molecular Genome Analysis, German Cancer
COMMENT sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFp762M2413) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES
            Location/Qualifiers
            source
            1..3243
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /map="19p13.3"
               /clone="DKFp762M2413"
               /tissue-type="melanoma (Memo cell line)"
               /clone_lib="762 (synonym: hmel2). Vector pSport1; host
               DH10B; sites NotI + SalI"
               /dev_stage="adult"
               1..3243
               /gene="DKFp762M2413"
               <1..1887
               /gene="DKFp762M2413"
               /note="unknown protein"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="CAD39039.1"
               /db_xref="GI:21740045"
               /translation="1MWANIETGEEERRLTFCHGLSNVLDDPKSAGVAFVIOEEDR
               FTGYMWCPTAEGSDGKIVSTOEKEIVQESLEPKVEYIARAOTRGKXAMAF
               LRPQOOLVLVLPALFLPSPNEBOIRLASAAVARNQPVYVEVYVWVNIYDI
               FYFPNORETSSAFSAFPMARPAKICRSPFLKSQGDDEPSPGDERCKPLK
               EETALTSGEWEVLAHRSKIWNEETKLVYFGTKDTPLEHLLVVSYEAGAEIVRLT
               TPQFSHSCMSQNFDMFVSHYSSTPCVHVYKLSGPDDELPHKQRFVASMMEAS
               CPEDVYPEIFHFHRTSDVLYGMITYKPHALOPGRKHPYLVFYGGPOVOLVNNSEFG
               IKYLRNLTASLGAVVVIDGSGCORGRFEGALKNOMQVEIEDOVEILOPVAEKY
               GFIDLSRAVIAHGSYSGFLSIMGLIHKPOVEKYAIACAPVTVMAYDCTERYMNP
               ENNOHCYEVAGSVALHVEKLPNEPNRLILHGFIDENVHFFHTNLFVLSOLIRAKPYOL
               QEYL"
               3199..3204
               /gene="DKFp762M2413"
               3226
               /gene="DKFp762M2413"
polyA_signal
polyA_site
BASE COUNT 731 a 977 c 911 g 624 t
ORIGIN
Query Match 70.9%; Score 1854.6; DB 9; Length 3243;
Best Local Similarity 99.7%; Pred. No. 4.6e-306;
Matches 1879; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```

```

Db 121 GAGTTGACCGCTTCACTGGGTACTGGTGGTGCCCAACAGCTCTCGGAAGTTCCAGAG 180
Qy 815 GGCCTCAAGACGCTGCGAATCTGTATGAGAACTCGATGATGCCAGAGTGGAGTCATT 874
Db 181 GGCTTCAAGACGCTGCGAATCTGTATGAGAACTCGATGATGCCAGAGTGGAGTCATT 240
Qy 875 CAGCTCCCTCTCTGGGCTGAGAAAGAAAGAGAGGAGCTGTGCTGATGCCAGAGCA 934
Db 241 CAGCTCCCTCTCTGGGCTGAGAAAGAAAGAGAGGAGCTGTGCTGATGCCAGAGCA 300
Qy 935 GGCAGCAAGATCCCAAGATTGCTTGAATGCTGCTAGTTCAGACTGACAGCAGAGGC 994
Db 301 GGCAGCAAGATCCCAAGATTGCTTGAATGCTGCTAGTTCAGACTGACAGCAGAGGC 360
Qy 995 AAGATGCTCTGACCCAGAGAGAGAGAGAGCTGTGACGCCCTTCAGCTGCTGCCAAG 1054
Db 361 AAGATGCTCTGACCCAGAGAGAGAGAGAGCTGTGACGCCCTTCAGCTGCTGCCAAG 420
Qy 1055 GTGAGTACATGCGCAGAGGCGCGGGTGGACCGGGATGGCAATAGCGCTGGGCGCATGTC 1114
Db 421 GTGAGTACATGCGCAGAGGCGCGGGTGGACCGGGATGGCAATAGCGCTGGGCGCATGTC 480
Qy 1115 CTGAGACCGGCCCCAGCAGTGGCTCCAGCTGCTCCCTCCCGCGGCGCTTTCATCCG 1174
Db 481 CTGAGACCGGCCCCAGCAGTGGCTCCAGCTGCTCCCTCCCGCGGCGCTTTCATCCG 540
Qy 1175 AGCAGAGAAATGAGAGAGCAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAG 1234
Db 541 AGCAGAGAAATGAGAGAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAG 600
Qy 1235 CCGTATGCTGCTACGAGAGAGAGAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAT 1294
Db 601 CCGTATGCTGCTACGAGAGAGAGAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAT 660
Qy 1295 CCGTATGCTGCTACGAGAGAGAGAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAT 1354
Db 661 CCGTATGCTGCTACGAGAGAGAGAGAGAGAGCTGACCTCTGACAGAGCTGTGCCAGAAATGTCAT 719
Qy 1355 ACCGGCTTCTGCCATTTGTACAAATGACGCGC-TTTTAAATCCAGAGGCTACGATTG 1413
Db 719 ACCGGCTTCTGCCATTTGTACAAATGACGCGC-TTTTAAATCCAGAGGCTACGATTG 779
Qy 1414 GAGTACCGCTTCCAGCGCCCGGGGAAGATGATTTAAGTCCCATTTAAGAAAGATTCG 1473
Db 779 GAGTACCGCTTCCAGCGCCCGGGGAAGATGATTTAAGTCCCATTTAAGAAAGATTCG 839
Qy 1474 TCTGACAGCGGTGAATGGAGAGTTTGGCGAGGACAGCGCTCCAAAGATCTGGTCAATGA 1533
Db 839 TCTGACAGCGGTGAATGGAGAGTTTGGCGAGGACAGCGCTCCAAAGATCTGGTCAATGA 899
Qy 1534 GGAGACCAAGCTGTGTACTTCCAGGGCACCAGAGACAGCGCGTGGAGACCACTCTA 1593
Db 899 GGAGACCAAGCTGTGTACTTCCAGGGCACCAGAGACAGCGCGTGGAGACCACTCTA 959
Qy 1594 CGTGTGAGCTATGAGCGCGCGCGGAGATGCTACGCTTACACAGCGCGGCTCTCCCA 1653
Db 959 CGTGTGAGCTATGAGCGCGCGCGGAGATGCTACGCTTACACAGCGCGGCTCTCCCA 1019
Qy 1654 TAGTGTCTCATGAGACCAAGCTTGCATGCTGTCAGCCTACACAGCAGCAGTGGAGAC 1713
Db 1019 TAGTGTCTCATGAGACCAAGCTTGCATGCTGTCAGCCTACACAGCAGCAGTGGAGAC 1079
Qy 1714 GCGCGCTGCTGACGCTTCAAGCTGAGCGCGCGGAGAGACAGCAGCAGCAGCAGCA 1773
Db 1080 GCGCGCTGCTGACGCTTCAAGCTGAGCGCGCGGAGAGACAGCAGCAGCAGCAGCA 1139
Qy 1774 GCGCGCTGCTGAGCTGACATGAGAGAGAGAGAGAGCGCGCGGATTTATGTCCTCC 1833
Db 1140 GCGCGCTGCTGAGCTGACATGAGAGAGAGAGAGAGCGCGCGGATTTATGTCCTCC 1199
Qy 1834 AGAGATCTTCATTTCCACACAGCGCTGAGATGTCGCTTACAGGCAATGATCTCAAGCC 1893
Db 1893 AGAGATCTTCATTTCCACACAGCGCTGAGATGTCGCTTACAGGCAATGATCTCAAGCC 1893

```

Db	1200	AGNATCTTCATTTCCACACGCGCTCGGATGTGGGGCTCTACGGCAGTATCTACAGGC	1259
QY	1894	CCAGCCTTTCAGCAGGAGAGAGACACCCACCGTCTCTTTGTATATGAGGCCCCCA	1953
Db	1260	CCACGCTTTTGCAGCCAGGAGAGACACCCCGCTCTTTGTATATGAGGCCCCCA	1319
QY	1954	GSTGAGCTGGGATATACCTCTTCAAGGCAATCAAGTACTGGGGCTCAACACTGTGC	2013
Db	1320	GTTGACGCTGGGATATACCTCTTCAAGGCAATCAAGTACTGGGGCTCAACACTGTGC	1379
QY	2014	CTCCCTGGGCTTCAGCGGTGTGATTTAGACGGCAGGGGCTCTGTCAAGCGGGCTTCG	2073
Db	1380	CTCCCTGGGCTTCAGCGGTGTGATTTAGACGGCAGGGGCTCTGTCAAGCGGGCTTCG	1439
QY	2074	GTTGCAAGGGGCGCTGAAAAACCAATAGGGCGAGGTGAGATCGAGACCAAGTGGAGGG	2133
Db	1440	GTTGCAAGGGGCGCTGAAAAACCAATAGGGCGAGGTGAGATCGAGACCAAGTGGAGGG	1499
QY	2134	CTTGAGATTTCGTGGCGGAGAAATATAGGCTTATCGACTGAGCCGAGTTGCCATCTATGG	2193
Db	1500	CTTGAGATTTCGTGGCGGAGAAATATAGGCTTATCGACTGAGCCGAGTTGCCATCTATGG	1559
QY	2194	CTGGGCTTACAGGGGGCTTCCTCGCTCATGAGGGGTAAATCCCAAGCCCCAGGTGTCAA	2253
Db	1560	CTGGGCTTACAGGGGGCTTCCTCGCTCATGAGGGGTAAATCCCAAGCCCCAGGTGTCAA	1619
QY	2254	GGTGGCCATTCGGGGGTGCCCCGCGTACCGTCTGATGGCTTACGACACAGGGTACACTGA	2313
Db	1620	GGTGGCCATTCGGGGGTGCCCCGCGTACCGTCTGATGGCTTACGACACAGGGTACACTGA	1679
QY	2314	GGGCTCATATGACATGCTCCCTGAGAACACAGCAGGCTATAGGGGGGTTCCTGTGGCCT	2373
Db	1680	GGGCTCATATGACATGCTCCCTGAGAACACAGCAGGCTATAGGGGGGTTCCTGTGGCCT	1739
QY	2374	GCACGTGAGAAAGTGGCCCAATGAGCCCAACCGCTTGCTTATCTCTCAAGGCTTCTGGA	2433
Db	1740	GCACGTGAGAAAGTGGCCCAATGAGCCCAACCGCTTGCTTATCTCTCAAGGCTTCTGGA	1799
QY	2434	CGAAAAAGTGACCTTTTTCACACAAACTTCTGTCTCCCAACTGATCCGAGCGGAA	2493
Db	1800	CGAAAAAGTGACCTTTTTCACACAAACTTCTGTCTCCCAACTGATCCGAGCGGAA	1859
QY	2494	ACCTTACCAAGCTCCAGATCTACCCC	2518
Db	1860	ACCTTACCAAGCTCCAGATCTACCCC	1884
RESULT 4			
AK054656			
LOCUS	AK054656	2261 bp	mRNA
DEFINITION	Homo sapiens CDNA FLJ30094 fls, clone BNGH41000034, weakly similar to DIERP1DYL PEPTIDASE IV (EC 3.4.14.5).		
ACCESSION	AK054656		
VERSION	AK054656.1	GI:16549242	
KEYWORDS	Oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens neurogliona cell, line:HA CDNA, clone_11b:BNGH41 clone:BNGH41000034.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Nishi, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K., and Iisaga, T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2261)		
AUTHORS	Iisaga, T., Otsuki, T., and Sugiyama, T.		

FEATURES	SOURCE	Location/Qualifiers
COMMENT		<p>Direct Submission</p> <p>Submitted (24-OCT-2001) Takao Isegai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hi.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.</p>
FEATURES	SOURCE	<p>Location/Qualifiers</p> <p>1. 2261</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="BNCH41000034"</p> <p>/cell_line="H4"</p> <p>/cell_type="neurogl1oma"</p> <p>/clone_lib="Neuro41"</p> <p>/note="Cloning vector: pME18SFL3"</p> <p>539..1477</p> <p>/note="unnamed protein product"</p> <p>/codon_start=1</p> <p>/protein_id="BAB70784.1"</p> <p>/db_xref="GI:16549243"</p> <p>/translation="MSQNFPMFVSHYSVSSTPPCVHYVYKLSGDDDLHKOPRFWMASMEASCPDYVPPEIFHFHRTSDYRLGMYIKPHALDPRGKHPIVLFVYGGPOVLVNSFGIKYLRNLNTASIGFVAVVYIDRGSORGLRFGALKNMGQVEIEDQVGLQFVAEYGFIDLSRVAIHGMSVGGFLSLMGLIHKIPQVEVYIACAPVTWMAVYDSYTERYMDPPEPNHCHYEKGSVALVHYEKIPNPNRNLILHGLDLENVHFFHNFPLVSQILRAGKPVOLQYIPNHRSHIRCPSESEHREVLTLHFLEHYL"</p>
BASE COUNT	470 a 704 c 640 g 447 t	
ORIGIN		
Query Match	56.5%	Score 1477.4; DB 9; Length 2261;
Best Local Similarity	99.9%	Pred. No. 8e-242;
Matches 1478; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
Oy	1126 CCAGAGATGGGCTCCAGCTCGTCGTCCTCCCTCCCGGCGCTGTTCAATCCGACACAGAGAA	1185
Db	1 CCAGAGATGGCTCCAGCTCGTCCTCTCTCCCGGCGCTGTTCAATCCGACACAGAGAA	60
Oy	1186 TGAGGAGCAGCGGCTAGCCTCTGCGCAGAGCTGTGCCAGAAATGTCCAGCGGTATGTGGT	1245
Db	61 TGAGGAGCAGCGGCTAGCCTCTGCGCAGAGCTGTGCCAGAAATGTCCAGCGGTATGTGGT	120
Oy	1246 GTACAGAGAGGTCAACCAAGCTGTGATCAATATGTCATATCTTCCTTCCTCCGCCA	1305
Db	121 GTACAGAGAGGTCAACCAAGCTGTGATCAATATGTCATATCTTCCTTCCTCCGCCA	180
Oy	1306 ATCAGAGGAGAGAGAGAGCTCTGCTTCTCCGCGCCCAATGAATGCAAGACCGGCTTCTG	1365
Db	181 ATCAGAGGAGAGAGAGAGCTCTGCTTCTCCGCGCCCAATGAATGCAAGACCGGCTTCTG	240
Oy	1366 CCATTGTGACAAAGTCACCGCGCTTTAAATCCAGGCGCTACGATGGAGTGAGCCCTT	1425
Db	241 CCATTGTGACAAAGTCACCGCGCTTTAAATCCAGGCGCTACGATGGAGTGAGCCCTT	300
Oy	1426 CAGCCCGGGGGAAGATGAATTTAATGTGCCCATTTAAGGAAGAGATTTGCTTGACACGGG	1485
Db	301 CAGCCCGGGGGAAGATGAATTTAATGTGCCCATTTAAGGAAGAGATTTGCTTGACACGGG	360
Oy	1486 TGAATGGGAGGTTTGGGAGGACAGGCTCCAAAGATCTGGGGTCATGAGGAGACCAACT	1545
Db	361 TGAATGGGAGGTTTGGGAGGACAGGCTCCAAAGATCTGGGGTCATGAGGAGACCAACT	420
Oy	1546 GGTGACTTCCAGGCGACCAAGAGACACCGCTGAGAGACCAACCTCTACGTGTACGTA	1605
Db	421 GGTGACTTCCAGGCGACCAAGAGACACCGCTGAGAGACCAACCTCTACGTGTACGTA	480
Oy	1606 TGAGGCGGCGGAGATGTACGCTACCAACGCGCGGCTTTCACATACGTGCTCAT	1665

Db 481 TGAGGGGGGGGGGAGATCTACGCTCACCAGCCCGGCTTCTCCATAGTCTCCAT 540
 QY 1666 GAGCAGACACTTGCATGTTGTCAGCACAACAGAGGGGTGAGCAGCGCCCTGCT 1725
 Db 541 GAGCAGACACTTGCATGTTGTCAGCACAACAGAGGGGTGAGCAGCGCCCTGCT 600
 QY 1726 GCAAGTCTACAGTGAAGCGGGCCGAGCAGAGAGCCCTGACAAAGCAGCCGCTCTG 1785
 Db 601 GCAAGTCTACAGTGAAGCGGGCCGAGCAGAGAGCCCTGACAAAGCAGCCGCTCTG 660
 QY 1786 GCGTACATATGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1845
 Db 661 GCGTACATATGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
 QY 1846 TTTCACACAGCGCTGAGTATGAGGCTTACAGCAGCAGCAGCAGCAGCAGCAGCAG 1905
 Db 721 TTTCACACAGCGCTGAGTATGAGGCTTACAGCAGCAGCAGCAGCAGCAGCAGCAG 780
 QY 1906 GCCAGGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1965
 Db 781 GCCAGGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 840
 QY 1966 GAATTAATCTCTTCAAAAGGATCAAGTACTTGGGCTCAACACTGGCTCCCTGGCTA 2025
 Db 841 GAATTAATCTCTTCAAAAGGATCAAGTACTTGGGCTCAACACTGGCTCCCTGGCTA 900
 QY 2026 CGCGCTGCTGTGATTAACAGCAGGAGGCTCTGACAGAGGCTTGGCTTCAAGGAGC 2085
 Db 901 CGCGCTGCTGTGATTAACAGCAGGAGGCTCTGACAGAGGCTTGGCTTCAAGGAGC 960
 QY 2086 CTTGAAAAACCAATGAGGAGGAGTGAAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 2145
 Db 961 CTTGAAAAACCAATGAGGAGGAGTGAAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 1020
 QY 2146 GCGCAGAGATGATGCTTCACTGACCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGG 2205
 Db 1021 GCGCAGAGATGATGCTTCACTGACCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGG 1080
 QY 2206 GGGCTTCTCTGCTCAATGAGGAGGAGTGAAGTGAAGAGCAGGAGGAGGAGGAGGAG 2265
 Db 1081 GGGCTTCTCTGCTCAATGAGGAGGAGTGAAGTGAAGAGCAGGAGGAGGAGGAGGAGG 1140
 QY 2266 GGGCTTCTCTGCTCAATGAGGAGGAGTGAAGTGAAGAGCAGGAGGAGGAGGAGGAG 2325
 Db 1141 GGTGCGCCCGGTACCGCTTGAAGGAGGAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 1200
 QY 2326 GGTGCGCCCGGTACCGCTTGAAGGAGGAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 2385
 Db 1201 GGTGCGCCCGGTACCGCTTGAAGGAGGAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 1260
 QY 2386 GGTGCGCCCGGTACCGCTTGAAGGAGGAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 2445
 Db 1261 GGTGCGCCCGGTACCGCTTGAAGGAGGAGTGAAGAGCAGGAGGAGGAGGAGGAGGAG 1320
 QY 2446 CTTTTCACACAAATCTCTGCTCCCACTGATCCGAGCAGGAGGAGGAGGAGGAGGAG 2505
 Db 1321 CTTTTCACACAAATCTCTGCTCCCACTGATCCGAGCAGGAGGAGGAGGAGGAGGAG 1380
 QY 2506 CCAAGATCTACCCCAAGCAGAGACACAGTATGCTGCGCCGAGTGGGCGAGCAGTATGA 2565
 Db 1381 CCAAGATCTACCCCAAGCAGAGACACAGTATGCTGCGCCGAGTGGGCGAGCAGTATGA 1440
 QY 2566 AGTCAAGTCTGCTGACTTCTTACAGGAATACCTGTAGC 2604
 Db 1441 AGTCAAGTCTGCTGACTTCTTACAGGAATACCTGTAGC 1479

RESULT 5
 BC000970 2546 bp mRNA linear PRI 12-JUL-2001
 DEFINITION Homo sapiens, clone IMAGE:3447394, mRNA, partial cds.
 ACCESSION BC000970 GI:12654298

KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2546)
 Strausberg, R.
 Direct Submission
 Submitted (16-NOV-2000) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-MGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalobpdm.tmc.edu
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

FEATURES
 source
 Location/Qualifiers
 1..2546
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3447394"
 /tissue_type="Placenta, choriocarcinoma"
 /clone_id="NH_MGC_10"
 /lab_host="DH10b"
 /note="Vector: pCMV-SPORT6"
 <1..1320
 /codon_start=1
 /product="Unknown (protein for IMAGE:3447394)"
 /protein_id="AAH00970.1"
 /db_xref="GI:12654298"
 /translation="DIFYPPOSEGEDELFLRANEKGTGFCHLYKVTAVLKSOGYDM
 SEPPSPGEDEFFKCPRIKEIETLSTGEWVLELRLHRSKIMVNEETLIVFOGCKRDPLEH
 LLYVSEAEAGEIYRLTPGSHSCMSQNDMEVSHSVSTPCVHYVRLSPDDP
 LHKOPRFWASMMEDASCPDPVPEPEIHFHTRSDVRLYGMIVYPHALDPEKRPVLE
 VVGGPOVQIVNNSEFKIKYRLNLTLASLGAVVAVIDGRSGORGLREGLAKNMQV
 EIEDVQEGLOFVAKEGFIIDLSVAIHGMSYGFSLIMGLIHRPOVEKVAIACAPVTV
 WMAVDGTEYRMDVPENNOHGEAGSVLAHVERKLPEPRHLILHGFEDENVHFFHT
 NPLVSOLIRAGKRYOLOIYRNEHRSIRCPESGSHYETVLLHFLQEVTL"

CDS
 556 a 795 c 714 g 481 t

BASE COUNT
 ORIGIN
 Query Match 50.5%; Score 1322; DB 9; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 2, 2e-215;
 Matches 1322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1283 GACATCTTCTATCCCTTCCCAATCAGAGGAGAGCAGCTGCTTCTCCGCGCC 1342
 Db 1 GACATCTTCTATCCCTTCCCAATCAGAGGAGAGCAGCTGCTTCTCCGCGCC 60
 QY 1343 AATGAATGCAAGACCGGCTTGCATTTGTACAAAGTACACCGCGCTTAAATCCAG 1402
 Db 61 AATGAATGCAAGACCGGCTTGCATTTGTACAAAGTACACCGCGCTTAAATCCAG 120
 QY 1403 GGCTACGATTGAGTGAAGCCCTTCAGCCCGGAGAGATGAATTAAGTCCCATTAAG 1462
 Db 121 GGCTACGATTGAGTGAAGCCCTTCAGCCCGGAGAGATGAATTAAGTCCCATTAAG 180

OY	1463	GAAGAGATTGCTCTGACCAACGGTGAATGGAGGTTTGGCAGGACGGCTCCAAATGTC	1522
Dp	181	GAAGAAGATTGCTCTGACCAACGGGTGAATGGAGGTTTGGCAGGACGGCTCCAAATGTC	240
OY	1523	TGGGTCAATGAGGAGACCAAGCTGTGTGTACTTCCAGGGGACCAAGACAGCGCGTGGAG	1582
Dp	241	TGGGTCAATGAGGAGACCAAGCTGTGTGTACTTCCAGGGGACCAAGACAGCGCGTGGAG	300
OY	1583	CACCACTCTACGTGTGTACGCTATAGAGGGGCGCGGAGATCTGACGCTCACACGCGCC	1642
Dp	301	CACCACTCTACGTGTGTACGCTATAGAGGGGCGCGGAGATCTGACGCTCACACGCGCC	360
OY	1643	GGCTTTCCCATTACTCTCTCCATGAGACCAAGAACTTGACATTTGTGTGACGCACTACAGC	1702
Dp	361	GGCTTTCCCATTACTCTCTCCATGAGACCAAGAACTTGACATTTGTGTGACGCACTACAGC	420
OY	1703	AGCGTGAAGACGCGCGCCCTCGTGCACAGTGTACAGCTGAGCGGCGCCGACGACGACGCC	1762
Dp	421	AGCGTGAAGACGCGCGCCCTCGTGCACAGTGTACAGCTGAGCGGCGCCGACGACGACGCC	480
OY	1763	CTGCACAAGACAGCGCGCCCTTCTGTGGGCTAGCATGTATGGAGGACGACGCTGCCCGCGAT	1822
Dp	481	CTGCACAAGACAGCGCGCCCTTCTGTGGGCTAGCATGTATGGAGGACGACGCTGCCCGCGAT	540
OY	1823	TATTTCTCTCCAGAGATCTTCCATTTCACACGCGCTCGGATGTGCGCTCTACGGCATG	1882
Dp	541	TATTTCTCTCCAGAGATCTTCCATTTCACACGCGCTCGGATGTGCGCTCTACGGCATG	600
OY	1883	ATCTACAAAGCCCCACGCGCTTTCACAGCCAGGAGAGACGCCACCGCTCTCTTTGTATAT	1942
Dp	601	ATCTACAAAGCCCCACGCGCTTTCACAGCCAGGAGAGACGCCACCGCTCTCTTTGTATAT	660
OY	1943	GGAGGCCCCCAGGAGGACGTGGGAGTAATCTCTTAAAGGATCAAGTACTGTGGGCTC	2002
Dp	661	GGAGGCCCCCAGGAGGACGTGGGAGTAATCTCTTAAAGGATCAAGTACTGTGGGCTC	720
OY	2003	AACACATGGGCTCTCCCTGGGCTACGGCGTGGTGTGATTGACGGCAGGGGCTCTGTTCAG	2062
Dp	721	AACACATGGGCTCTCCCTGGGCTACGGCGTGGTGTGATTGACGGCAGGGGCTCTGTTCAG	780
OY	2063	CGAGGCGCTTGCGTTCCGAAGGGGCGCCGTGAAGAAACCAATGGGCGACAGTGCAGATCGAGAC	2122
Dp	781	CGAGGCGCTTGCGTTCCGAAGGGGCGCCGTGAAGAAACCAATGGGCGACAGTGCAGATCGAGAC	840
OY	2123	CAGGTGAGAGGGCTGCAGTTCGTGTGGCGGAGAGTATAGCTTATGACACTGACGCGAATT	2182
Dp	841	CAGGTGAGAGGGCTGCAGTTCGTGTGGCGGAGAGTATAGCTTATGACACTGACGCGAATT	900
OY	2183	GGCATTCATGGCTGTGCTCTACGGGGGCTTCTCTGTGCTATGGGCTTAATCCACAGGCC	2242
Dp	901	GGCATTCATGGCTGTGCTCTACGGGGGCTTCTCTGTGCTATGGGCTTAATCCACAGGCC	960
OY	2243	CAGGTGTTCAAGGGTGGGCTATCGGGGGGTGCCCGGTGACCGTGTGGATNGGCGTACGACACA	2302
Dp	961	CAGGTGTTCAAGGGTGGGCTATCGGGGGGTGCCCGGTGACCGTGTGGATNGGCGTACGACACA	1020
OY	2303	GGGTACACTGAGGCGCTACATGAGAGTCCCTGTGAAACCAACACAGCGGCTATGAGGGGGGT	2362
Dp	1021	GGGTACACTGAGGCGCTACATGAGAGTCCCTGTGAAACCAACACAGCGGCTATGAGGGGGGT	1080
OY	2363	TTCGTTGGCCCTCGCACGTGGAGAAAGCTGGCCCATGAGCCCAACCGCTTGTGCTTATCTCTCAC	2422
Dp	1081	TTCGTTGGCCCTCGCACGTGGAGAAAGCTGGCCCATGAGCCCAACCGCTTGTGCTTATCTCTCAC	1140
OY	2423	GGCTTCTTGAGACGAAAGCTGCACTTTTTCACACAAATCTCTGTCTCCCAACTGATC	2482
Dp	1141	GGCTTCTTGAGACGAAAGCTGCACTTTTTCACACAAATCTCTGTCTCCCAACTGATC	1200
OY	2483	CGAGCAGGAGAACTTACAGAGTTCAGATCTACCCAACGAGAGACAGATATTTGCTGTC	2542
Dp	1201	CGAGCAGGAGAACTTACAGAGTTCAGATCTACCCAACGAGAGACAGATATTTGCTGTC	1260

QY	2543	CCGAGTGGGGGACGACCATATGAAGTACAGGTGCTGACACTTCTACAGAAACTCTGCA	2602
Db	1261	CCCGAGTGGGGGACGACCATATGAAGTACAGGTGCTGACACTTCTACAGAAACTCTGCA	1320
QY	2603	GC 2604	
Db	1321	GC 1322	
RESULT 6	AX354795	2649 bp	DNA
LOCUS	AX354795		Linear
DEFINITION	Sequence 3 from Patent WO0179473.		PAT 06-FEB-2002
ACCESSION	AX354795		
VERSION	AX354795.1		GI:18619528
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1		
JOURNAL	Meyers, R.A. and Williamson, M.		
FEATURES	21553, a human prolyl oligopeptidase family member and uses thereof		
	Patent: WO 0179473-A 3 25-OCT-2001;		
	Millennium Pharmaceuticals, Inc. (US)		
	Location/Qualifiers		
	1..2649		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	803 a	514 c	586 g
ORIGIN		746 t	
Query Match	33.0%;	Score 862.6;	DB 6;
Best Local Similarity	59.4%;	Pred. No. 3.7e-137;	
Matches 1500;	Conservative	0;	Mismatches 1019; Indels 6; Gaps 2;
QY	80	TTTCAGGTGCAGAGACACTGTGGGAGCGGGCTCCGGAGCATATCCACGCGCAG	139
Db	103	TTTATGTGTGACGGGTATTCCTGGAGTCAGCTTAAAGACTCTTGGCGATACGAGAAA	162
QY	140	TACTGGGGCTATTTGTCAACAAGCGCGCCACGACTTCCAGTTTGTGCAAGAAGCGAT	199
Db	163	TATCATGGCTATCATGTGGCTTAAGGACACACATGATTTTCATGTGTAAGAGGATGAT	222
QY	200	GAGTGTGGGCCCCCAGTCCACCGCCTCTACTACTGTGGGAATGCCAATATGGCAGCGAGAG	259
Db	223	CCAGATGGACCTCATTTACAGACAGAAATCTATTACCTTCCATCTCTGTGTGAGAACAGAGAA	282
QY	260	AACCTCCTCTCTACTCTGAGATTCGCAAGAAGTCCGAAAGAGGCTCTGCTCTCTG	319
Db	283	AATPACACTGTTTATTATTCGGAATTTCCCAAAACTATACATAAGACAGCAAGTATTATGCTC	342
QY	320	TCTTGGAGACAGATCTGGATATTTCCAGGCGCCACGCCCCACATGGGGTCTACTCTGG	379
Db	343	TCTTGGAGCCCTCTTTTGGATCTTTTTCAGGCGAACACTGAGACTATGGAATGATTTCTGGA	402
QY	380	GAGGAGGAGCTCTAGGAGGAGGAAAGCGCTTGGGGGTCTTCCGGATCATCCCTCAAGAC	439
Db	403	GAGAGAGAACTATTAAAGAGAAAGAAAGCCATTGGAACAGTGGAAATGCTTCTTAACGAT	462
QY	440	TTTCACAGAGAGAGTGGCTCTTCCCTTCCAGGCGCCACAGCAGCTCTCCACACTGTGCG	499
Db	463	TATTCACCAAGAGAGTGGAAATTTCTGTTTCAAGCGCGGATAGTGGAAATTTATCACGTTAAA	522
QY	500	GACGGCGCAGAGACGCGCTTCATGCTGCCCTATGAAACGCGTGGAAATCAAGACCCAG	559
Db	523	GATGAGGGGCCACAGAGATTTTACGCAACAACCTTTAAAGCCCAATCTAGTGGAAACTAGT	582
QY	560	TGCTCAAGGGCCCCGGATGGACCCCAAAATCTGCCCCGCGGACCCCTGCTTCTTCTCTTC	619
Db	583	TGTCCCAACATATGAGATGGATGAAATTTATTCGCTGATCCAGACTGATTCCTTT	642
QY	620	ATCATTAACAGGAGCTCTGGGTGCGCAACATCGAGACAGGCGAGAGACGGCGGCTGACC	679

[illegible]

Query Match	Best Local Similarity	Score	DB 6:	Length	3143:
Matches 1500:	Conservative	0:	Mismatches 1019:	Indels	Gaps
80	TTCCAGGTGCAACAAACATCTCGTGGAGCGGGCCCGGAGCATATCCACAGCGCCGCAAG	139	11	11	11
Db	331	TTTTATGTTGAGCGGATATTCCTGAGAGTACAGCTTAAAAAGCTCTCGCGATACCGAAAA	390	11	11
Qy	140	TACTCGGGGCGCTATTGTCAACAAGGCGGCCACGACTTCCAGTTTGTGCAGAGAAGCAT	199	11	11
Db	391	TATCATGTGGTACATGATGAGCTAAGGACACACATGATTTGATGTTGTGAAGAGGAATGAT	450	11	11
Qy	200	GAGTGTGGGCGCCCATCTCCACCGCCTCTACTACTGAGTAATGCCATATGGACCGGAGAG	259	11	11
Db	451	CCAGATGGACCTCATTCAGACACGAATCTATTACCTTGGCATCTGTGATGAGAACAGAGA	510	11	11
Qy	260	AACCTCCCTCCCTACTACGTGAGATTTCCCAAGAGGTCGCGAAGAGAGGCTCTGCTGCTCG	319	11	11
Db	511	AATTAACACTGTTTATTATTCTGAAATTTCCCAAACTATATCATAGACAGCACTCTTAATGCTC	570	11	11
Qy	320	TCCCTGAAGACAGATCTGTGATCTTTCAGGCGCACGCCCCACCATGAGGGGTCTACTCTCG	379	11	11
Db	571	TCTTGAAGCGCTCTTTTGGATCTTTTTCAGGCAACCTGAGACTATGAAATGTAATTCCTGA	630	11	11
Qy	380	GAGGAGGAGACTCTGAGGAGCGGAAAGCGCTGGGGGCTTTGGGCATCACTTCCTACGAC	439	11	11
Db	631	GAAAGAGAAGCTTTTAAGAGAAAAAGAACGCACTTGGACAGTGTGGAATGCTCTTTACGAT	690	11	11
Qy	440	TTCCACAGGAGAGTGGCCCTCTCCCTTCGACGGCAGCAACAGGCTCTTCACAGTGTGC	499	11	11
Db	691	TATTCACCAAGAGATGGAACTTTCTGTTCAAGCGGGATGATGGAAATTTATCACGTAATA	750	11	11
Qy	500	GACGGCGGCAAGACGCGCTTCATGTTGTCCTCATGAAACCGCTGGAATATCAAGACCCAG	559	11	11
Db	751	GATGGAGGGGCCACAAAGATTTACGCAACAACCTTTAAGGCCCAATCTATGGAAACATGAT	810	11	11
Qy	560	TGCTAGAGGGCCCCGGATGAGACCCCAAAATCTGCCCTCGGACCTGCTCTTCTCTTC	619	11	11
Db	811	TGTCCCAACATCATGATGATGCCAAATATATCCCTGTGATCCAGACTGAGATTCCTTT	870	11	11
Qy	620	ATCATTAACAGGAGCTGTGGGTGGGCAACATTCGAGACAGGAGGAGGAGGGGGGCGAC	679	11	11
Db	871	ATACATAGCAACGATATTTTGGATATCTTAACATCGTAAACGAGAGAAAGAGACTCACT	930	11	11
Qy	680	TTCTCGACCAAGATTTATCCAAATGCTCGATGAGTGAACCCCAATCTGCGGGTGTGGCCAC	739	11	11
Db	931	TATGTGCAACAATGAGACTAGCCCAACATGGAAGAAAGATGCCAATGCACTGTGATCGCTAC	990	11	11
Qy	740	TTTGCTCATACAGAGAGATTGACCGCTTCACTGGGTACTGTGTGTGCCCCACAGCTTC	799	11	11
Db	991	TTTGTCTTCCCAAGAAATTTGATGATATTCCTGGCTATTGGTGTGTCACAAAGCCTGAA	1050	11	11
Qy	800	TGGGAAGGTTACAGAGGGCCTCAAGACGCTGGCAATCTGTATGAGGAAGTGCATAGTGC	859	11	11
Db	1051	AC-- --AAGTCCAGTGGTGTGTAATTTCTTAAGATTTCTATATGAGAAAGAAATATGAAATCT	1107	11	11

QY 860 GAGGTGAGGTCATTCACGTCCCTCTCTGCGCTAGAGAAAGAGAGAGACTCGTAT 919
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1108 GAGGTGGAATTTATTCATGTATCATATCCCTATGTGTGAAACAAGAGAGGAGATTCATTC 1167
QY 920 CGGTACCCCGAGGAGGAGCAGAGAAATCCCAAGATTGCTTTGAAACTGGCTGAGTTCCAG 979
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1168 CGTTATCTCTAAACAGGTACAGCAAAATCTAAAGTCACTTTTAAAGATGAGAAATATATG 1227
QY 980 ACTGACAGCCAGGAGCATGTCTCGACCCAGAGAGAGAGGAGTGGTGGAGCCCTTGAGC 1039
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1228 ATTGATCTCTAAGAGAGATCATATGATGTATGATTAAGGACTAATTAACCTTTTGAG 1287
QY 1040 TCGCTGTTCGCCAAGGTGAGATACATCGCCAGGCGCGGTGAGACCCGCGATGCGAATAC 1099
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1288 ATTCTATTTGAAGAGATTGAATATATTTGCGAGACCTGGATGAGACTCTCGAGGAAAAATAT 1347
QY 1100 GCGTGGGCCATGTCTCTGAGACCGGCCAGCAGTGGCTCCAGCTGCTGCTCTCCGCCCG 1159
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1348 GCTTGGTCCATCTACTACATGCTGCCAAGCTGCTGAGATGATGTGATCTTCACT 1407
QY 1160 GCCCTGTTCATCCCGAGCAGACAGATAGAGAGAGCGGCTAGCCTCTGCCAGAGCTGTG 1219
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1408 GAATTAATTTATCCCATGAGAAGATGATGTATGAAAGCGCAGACATTCATTGAGTCACTG 1467
QY 1220 CCCAGGATGTCCAGCCGTATGTGTGTACAGAGAGGTCCACCAAGCTGTGATCAATGTT 1279
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1468 CCGTATCTGTGACGCCACTAATATATCTATGAAAGAACACAGACATCTGATTAATATTC 1527
QY 1280 CATGACATCTTCTATCCCTTCCCCCATCAGAGGAGAGAGACGAGCTCTGCTTCTCCGC 1339
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1528 CATGACATCTTCTATGTTTTCCTCCCAAGTACAGAGAG--GAAATTAAGTTTATTTT 1584
QY 1340 GCCAATTAATGCAAGACCGGCTTCTGCCATTTGTACAAGTCAACGCGCTTTAAATTC 1399
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1585 GCGCTCTAATGCAAAACAGGTTTCCGTCATTTATACAAATTTACATCTATTTTAAAGAA 1644
QY 1400 CAGGGCATGATTGAGTGAAGCCCTTCAGCCCGGAGAGATTAATTAAGTCCCATTT 1459
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1645 AGCAAAATATAAAGATCAGTGTGTGCGCTGCTCCAAAGTATTTCAAGTGTCTATTC 1704
QY 1460 AAGGAGAGATGTCTCTGACACAGGGTGAATGGAGGTTTGGGAGAGCACGCTCCAG 1519
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1705 AAGAGAGATGATGAAATTCAGTGTGAATGGAGATCTTGGCGCGATGATTCAT 1764
QY 1520 ATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGACCAAGACACGCGCTG 1579
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1765 ATCCAAATTTGATGAAGTCAGAGGCTGTATATTTTGAAGCACCAAGACTCCCTTTA 1824
QY 1580 GAGCACCACCTTACGTGTGAGCTATAGAGCGGCGGAGATCGTACGCTTACCACG 1639
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1825 GAGCATCACCTGTACGTAGTACGTATTAATTCGTGAGAGGTGACAAAGGCTACGTGAC 1884
QY 1640 CCGGCGCTTCCCATAGCTGCTCATGAGCAGAACTTCAGATGTTGCTGACGCACTAC 1699
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1885 CGTGGCTACTCAATTTCTGTGATCATGTCAGTGTGATCTTTATTAAGTAATAT 1944
QY 1700 AGCAGCTGAGACAGCGCGCTGCGTCAAGCTTCAAGCTGAGCGGCGCCGAGCAGAC 1759
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 1945 AGTAAACCAAGAAATCCACACTGTGTCTCTTTACAAAGTATCAATCTCTGAAGATGAC 2004
QY 1760 CCGCTGACAGAGAGCCCGCTTCTGCGGTAGCATGATGAGAGCAGCAGCTGCGCCCG 1819
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2005 CCAACTTGCAAAACAAAGAAATTTTGGCCACACATTTTGGATTTCAGAGGTCCTTCT 2064
QY 1820 GATTATGTCTCTCAAGATCTTCCATTTCCACACGCGCTCGATGTGCGGCTTACGGC 1879
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2065 GACTATATCTCTCCAGAAATTTCTTCTTTGAAGTACTACTGATTTACATTTGTATGG 2124
QY 1880 ATGATCTACAGCCCGCGCTTGCAGCCAGGAGAAAGAACCCACCGTCTCTTGTGA 1939
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2125 ATGCTCTACAAGCTCTATGATCTACAGCCTGGAAAGAAATATCTTACTGTGCTTCTATA 2184

QY 1940 TATGAGAGCCCCAGGTGACGTGCTGTAATTAATCTTTCAAAGCATCAAGTACTGCGG 1999
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2185 TATGTGTGCTCTCAGGTGCTGATGTAATATATCGTTTAAAGAGTCAAGATTTTCCG 2244
QY 2000 CTCAACACTGCGCTCCCGGCTACGCCGTGGTTGTGATTTAGAGGCAAGGGGCTCCTGT 2059
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2245 TTGATATCCCTAGCGCTTCAAGTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2304
QY 2060 CAGGAGAGGCTTCGTTTCGAAGGGCCCTGAAAAACAAATGGGCGCAGTGGAGATGAG 2119
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2305 CACCGAGGCTTAATTTTGAAGGGCGCTTTAAATATTAATTTGGTCAAAATATGATGAC 2364
QY 2120 GACCAAGTGAAGGGCTGCAATTCGTGCGGAGAAATGATGCTTCATGACTGAGCGA 2179
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2365 GATCAGGTGGAAGAGACTCAATATCTATCTTCGATATATATTTCAATGACTTATGATCT 2424
QY 2180 GTTCCATTCATGGCTGTCTACGGGGGCTTCTCTCGCTCATATGGGCTTAATCCAAAG 2239
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2425 GTGGGCAATCCAGCGCTGTCTATGAGAGATACCTCTCCGTGATGGCAATTAATGACAGG 2484
QY 2240 CCGCAGGTGTTCAAGGTGCGCATGCGCGGTGCGCCGCTCACCGCTGTGATGCTTACGAC 2299
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2485 TCGATATCTTCAAGGTTTGTATGCTGTGGGCGCCACACTCATCTGTGATCTTCTATGAT 2544
QY 2300 ACAGGTACACTGAGCGCTACATGACGCTCCGTGAGAACAAACAGCAGCGCTATGAGCG 2359
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2545 ACAGGATACACGGAACCTTATGAGGTACACCTGACCCAGCATGATGACAGGCTTATCTTA 2604
QY 2360 GATTTCGGGCGCTGACGCTGAGAGAACTGCCAATAGGCCCAACGCTGTGATATCTTC 2419
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2605 GATCTGTGCGCATGCAAGCAGAAAGTTCCCTCTACCAAAATGCTTACGTGCTCTTA 2664
QY 2420 CACGGCTTCCTGACGAAGAACTGTCATCTTTCACACAAACTTCTGCTCCAACTG 2479
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2665 CATGTTCTCTGATGATGAAATGTCATTTTGCACATACGATATATACGAGTTTGA 2724
QY 2480 ATCCGACAGGGAACCTTACAGCTCCAGATTCACCCCAAGCAGACAGCATGATTCG 2539
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2725 GTGAGGGCTGGAAGCCATGTATGATTTACAGATCATCTCTGAGACACAGCATAGA 2784
QY 2540 TGCCCCGAGTGGGCGGAGCATATGAAGTCAAGTGTGCTGACCTTTCACAGGAATATCTC 2599
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2785 GTTCTGAATGGGAGAACATTTATGAACGTGATCTTTGCACTTCTCAAGAAACCTT 2844
QY 2600 TGAGC 2604
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2845 GGAATC 2849

RESULT 9
AF221634 3127 bp mRNA linear PRI 05-NOV-2000
LOCUS
DEFINITION
Homo sapiens dipeptidyl peptidase 8 (DPP8) mRNA, complete cds.
ACCESSION
AF221634
VERSION
AF221634.1 GI:11095187
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3127)
Abbott,C.A., Yu,D.M., Woollett,E., Sutherland,G.R., McCaughan,G.W.
and Gorrell,M.D.
Cloning, expression and chromosomal localization of a novel human
dipeptidyl peptidase (DPP) IV homolog. DPP8
Eur. J. Biochem. 267 (20), 6140-6150 (2000)
JOURNAL
MEDLINE
20467194
PUBMED
2 (bases 1 to 3127)
REFERENCE
Abbott,C.A., Yu,D., McCaughan,G.W. and Gorrell,M.D.
Direct Submission
Submitted (06-JAN-2000) A.W. Morrow Gastroenterology and Liver
Centre, Centenary Institute of Cell Biology and Cancer Medicine,

FEATURES		Locked	Bsg No.6,	Newtown,	Sydney,	NSW 2042,	Australia
Source	Location/Qualifiers						
gene	1. .3127 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="15" /map="15q22" 1. .3127 /gene="Dpp8" 214. .2862 /note="Dpp8" /note="serine protease; member of prolyl oligopeptidase 59b family belonging to enzyme clan SC; alternatively spliced forms of this mRNA are found in AF221635, AF221636 and AF221637"						
CDS	/codon_start=1 /product="dipeptidyl peptidase 8" /protein_id="AA029766.1" /db_xref="GI:11095188" /translation="MAAMETEQLGVEIFETADCEENIESODRPLEPPYVERYSWSQ LKLTLADTRKYHGYMMAKAPDHFEMVRKNDPDGSHSRILYLAAGSRENTLFLYSID PKTINRAVAILMSKPLDLDFEATLDVAGMSREBELREKRITGVLAISYDQSGSG TELEAGSGIYHVKDDGGPOGFTOOLPLNLTVEECSPNIRMDPKICPADPWIATHSN DITLSNITREPERLTYVHNELAMMEDBASAGATVYLOEPRPISGYWMCRAETTP PSGKILRILEYENDESVELIHTSPLETRRADSRFRPKTGIANRVFKMSLEIMT DAERIIDLVDKELIQPELLEFGEVYIARAGMPEKKAASILLDRSORLOIVLIS PELPIPEVDDMERORLIESVPDSVTPLIAYEETDIIWNIHDLHFPPSHHEEIEPE IFAECKTGFHRLIKYISILKESKYRSSGLPAPSPCKIKEIYAITSCEWELGR HGSNIQYDEVRLVLYFEGTQDSPLEHLIYVSVYNPGEVRLPDGQSHSCISQHD FEIKRYSNOKNPHGVSLKLSSPEDDPCKTKERWATILDSAGLPDTPPEITSFES TVVTLTGMLKLRPHLOPGRKYPTLVLEYGPOVLVNNRKGVYERLNTLASLGVV VVVDNRSGCHRGKLFEGAEFYKMGQLEIDQVQGLQYLAISRYDILDLRVHGSYV LSEKSLMALMRSDIFRYALIGAVPTLMIIFYDTGYTERMGHPQONBGQYLSVAQ AEKPESEPNRLILHGFLEDENVHPAHSILSLFLVRAKGRPYDLQIYPERHSINVPES GEHELHLHLHLONLGSRILAKLVI"						
BASE COUNT	943 a 637 c 706 g 841 t						
ORIGIN							
Query Match	32.9%; score 861; DB 9; Length 3127;						
Best Local Similarity	59.4%; Pred. No.6,7e-137;						
Matches 1499; Conservative	0; Mismatches 1020; Indels 6; Gaps 2;						
80	TTTCAGGTGCAGAAACACTCGTGGAGGGGCTCCGGAGCATCATCCACGGCAGCCGCAAG 139						
316	TTTATGTGAGACGGGTATTCTCGTAGTCAAGCTTAAAGATCGCTTGGCATACGAGAAA 375						
140	TACTCGGGCTCATTTGTCAACAAGCGGCCGCCACGACTTCCAGTTGTGCGAAGAAGCGAT 199						
376	TATCATGGCTACATATGCTAGGTCACACACATGATTTCTATGTTTGTGAAGGGATGAT 435						
200	GAGTGTGGGCCCATCCACCGCCCTTACTACTCGTGGGAATCCGATATGAGCGAGCCGAGAG 259						
436	CCAGATGGACCTCATTCACAGACGAATCTATTACCTTGGCATGTCTGGTGAGAACGAGAA 495						
260	AACTCCCTCTCTACTCTGAGATTCCCAAGAAGTCCGGAAAAGAGCTCTGCTGCTCTCG 319						
436	AATACACGTGTTTATTCTGGAATTTCCCAAAATATCAATATAGAGCAGAGCTTATATGCTC 555						
320	TCTCTGGAAGACATCTGTGATATTTCACAGGCCACGCCCCACCATAGGGGTCTACTCTCGG 379						
556	TCTTGGAAACCTCTTTTGGATCTTTTTCAGGCGAACACATGAGATATGATATTTCTGGA 615						
380	GAGGAGGAGCTCTCAGGGAGCGGAAAGCCCTGGGGGTCTTTGGCATCAGCTCTTACAGC 439						
616	GAAAGAAAGACTTATTAAGAGAAAGAAAACCAATGGACATAGTGGAAATTCCTTTACGAT 675						
440	TTTCACAGGAGAGTGGCCCTTCTCTTTCACAGGCCACGACACAGCCCTTCCACTGTGCGC 499						
676	TATTCACCAAGGAAGTGAACATTTCTGTTTCAAGCCGGTAAGTGAATTTATCAGCTAAAA 735						
500	GACGGCGGACAGAACGGCTCATGTGTGCCCTTATGAACCGCTGGAATATCAAGACCCAG 559						
736	GATTGAGGGGCCCAAGGATTTTACGACACACCTTTTAAGGCCCAATCTATGTGCAAACTGAT 795						

OY	560	TGTCACGAGGCGCCGGATGAGACCCCAAAATCTGCGCTGCCAGACCCGCTCTCTCTCCCTC	619
Db	796	TGTCCCAACATACATGAGGATGATCCAAAATTATGCCCCGGCTGATCCAGACTGATTCGCTTT	855
OY	620	ATCATATACAGCGCACTGCTGGTGGGCAACATTCGAGACACAGCGGAGAGCGCGCGCTGCAC	679
Db	856	ATACATATGACAACGATATTTTGGATATCTCAACATGTAACGAGAAAGAAAGAGACTGACT	915
OY	660	TTTCGCCACCAAGGTTATATCCATATGTCCTGATGACCCCAAGTCGCGGTGTGGCACCC	739
Db	916	TATGTGACACATGAGCTGAGGCCAACATGGAAGAAAGAGCCAGATCTAGCTGAGTGCCTACC	975
OY	740	TTTCGTATACGAGAAGATTCGACGCGTTCACTAGGGTACTGGGTGGCGCCACACACTCC	799
Db	976	TTTGTCTCCAGAGAAGATTGATAGATATATCTGGCTATATGGTGTGTGCCAAAAGCTGAA	1035
OY	800	TGGGAAGTTCAGAGGGCCCTCAAGACGCTGCGAATCTGTATAGAGAAAGATGATGATGTC	859
Db	1036	AC---AACTCCCACTGGTGGTAAAAATTTCTTAAGTAATCTATATGAGAAATAATGATGAACT	1092
OY	860	GAGGTGAGAGCTATTCACGTCGCCCTCTCCGGCCTAGAAAGAAAGAGACGCACTGAT	919
Db	1093	GAGGTGAGAAATTATTCATATGTTATCATCCCTATGTTGAAACAGAGGGCAATTCATC	1152
OY	920	CGGTACCCCAAGAGCAGGCAACAGAAATCCCAAGATTGCCCTTAACATGCTAGTTCAG	979
Db	1153	CGTATTCCTAAACAGGTTACAGCAAAATCCTAAAGTCACCTTTTAAGATGTGCAGAAATATG	1212
OY	980	ACTGACACCCAGGGCAAGATCTCTGCAACCAGAGAGAGAGCTGGTGCACTCCCTTCAGC	1039
Db	1213	ATTATATGCTGAAGAGAAAGATCATAGATATGTCATGATTAAGAAACTATATCAACCTTTTAG	1272
OY	1040	TCGCGTTCGCCAAGGTGGAGTACATCGCCAGGAGCGGGGAGCCCGGGATGGCAAAATAC	1099
Db	1273	ATTCTATTTGAGAGAGTTGAAATATTTGCCAGAGCTGAGATGCACTCTGAGGGAAATAT	1333
OY	1100	GCTTGCGCCATGTTCTTGAGACGGGCCCCAGACAGATGGCTCCAGCTGCTCCTCCGCCCG	1159
Db	1333	GCTTGCTCATCTCACTAGATAGATCCCTCCAGACTCGCCTACAGATGATGTTGATCTCACCT	1392
OY	1160	GCCCTGTTACCTCCGAGACACAGAAATGAGAGAGACAGGCTGACCTCTGCCAGACTCTC	1219
Db	1393	GAATATTTATTTATCCAGTAGAAGATGATGTTATGGAAGAACAGACACTCATATGATCACTG	1452
OY	1220	CCCAAGATGTCACCGCTATGTGTGTATCAGAGAGTCAACCAAGCTGTGATCAATGTT	1279
Db	1453	CTGTATCTCTGAGCGCACTAATATCTATGAAAGAAACAAGACATCTGGATTAATATC	1512
OY	1280	CATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGAGAGAGCTCTGCTTCTCCG	1339
Db	1513	CATGCACTCTTCACTGTTTTTCCCAAAAGTCAAGAAAG---GAATATGAGTTATTTTTT	1569
OY	1340	GCCATATGAATGACAGACCGGCTCTGCACTTTTACAAAGTCAACGGCTTTTAAATCC	1399
Db	1570	GCTCTGATATCAAAACAGGTTTCCGTCATTTATCAAAATTTACATCTATTTTAAAGAA	1628
OY	1400	CAGGCTACGATTGAGTGAAGCCCTTCAGCCCGCGGGAAGATGAATTTAAGTCCCAATT	1459
Db	1630	AGCAAAATATTAACGATCCAGTGGGGCTCTGCTCTCAAGATGATTTCAAGTGTCTATC	1689
OY	1460	AAGGAAGATGCTCTGACCAGCGGTGATGGGAGGTTTGGCAGAGCGAGCTCCAG	1519
Db	1690	AAAGGGGATAGCAATTCACAGGGGTGAATGGAAAGTTCTTGGCCGGATGATCTAAT	1749
OY	1520	ATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGACCAAGAAAGACGCCGTG	1579
Db	1750	ATCCAAGTTGATGAAGTCAGAGGCTGTATATTTTGAAGGCCACCAAAAGATCCCTTTTA	1809
OY	1580	GAGCAACACCTCTACGTGTGCTACGTATGAGAGCGCGCGGGAATCGTACGCTCACACG	1639
Db	1810	GAGCATCACTGTACGTATGATTTAGTAAATCTGAGAGAGGTGACAAGGCTGTACTAC	1869

1640 CCCGCTTCCCATAGCTGCTCCATGAGCAGAACTTGCAGATGTTCTGACCACTAC 1699
1870 CGGGGCTACTCATATCTTGTGCTCATGACAGCTGATCTTCTTATTAAGTAT 1929
1700 AGAGGCTAGAGCGCGGCTGCGTACGCTTACAGCTGAGCGCGCGGAGCAGC 1759
1930 AGTAACCAAGAGATCCACAGCTGTGTGCTTACAGCTATCAAGCTCTGAAGTAC 1989
1760 CCCCTCAGCAGAGCGCGGCTTGGGCTAGCATGATGAGGAGCAGAGCTCCCGC 1819
1990 CCACACTGCAAAACAAAGAAATTTTGGCCACCATTTTGGATTCACAGAGTCTTCC 2049
1820 GATTATGTTCTCCAGAGATCTTCAATTCACAGCGCTGAGATGCGGCTCTACGCG 1879
2050 GACTATACCTCTCCAGAAATTTTCTCTTTGAAGAAGTACTGATTTACATTTATMG 2109
1880 ATGATCTAAGCCCGCCTTGCAGCCGAGGAAAGACCCACCGCTCTTGTGA 1939
2110 ATGCTCTAAGGCTCATATGCTACAGCCGGAAGAAATATCTACTGCTGCTTCA 2169
1940 TATGAGGCGCCGAGGCTGAGTGAATTAATCTCTTCAAGGCAATCAAGTACTGCG 1999
2170 TATGAGGCTCTCAGAGTGTGAGTAATCGGTTTAAAGAGTCAAGTATTTCCG 2229
2000 CTCACACACTGCGCTCCCTGGGCTACGCCGCTGTTGTATGAGCGGAGGCTCTGT 2059
2230 TTGAATACCTGAGCTCTCTAGAGTATGTGTGTAGTATGACAAACAGGGGATCTGT 2289
2060 CAGCGAGGCTTCCGTTCAAGGGGCTGAAACCAATGGGCCAGAGTGAAGTACGAG 2119
2290 CAGCGAGGCTTAAATTTGAAGGCGCTTAAATTAATTAATGAGTGAATGAGTAC 2249
2120 GAGCAGAGGAGGCGCTGAGTCTGAGCGGAGAGATGAGTCAATGAGTCAAGTCA 2179
2350 GATCAGAGTGAAGGATCCAAATATCTAGCTTCGATATGATTTCAATGACTAGATCG 2409
2180 GTTCCATCATGAGTGTGCTCTACGGGCTTCTCTGCTATGAGGCTATTCACAG 2239
2410 GTGGGATCTCAGGCTGTGCTATGAGGATACCTCTCTGATGAGTATGACAGAG 2469
2240 CCCAGGTTCAAGGTGCGATGCGGGGCGCGGCTACCGCTGAGTGGCTGACGAG 2299
2470 TTAGATATCTTCAAGGTTGCTGATGCTGAGGCGCCAGTCTGTGAGTCTTATGAT 2329
2300 ACAGGCTACTGAGCTCATATGAGTGCCTCTGAGAACACAGCAGGCTATGAGCG 2359
2530 ACAGGATACAGGAGTATATGAGTGCACCGTACACAGAGTACAGAGGCTATTA 2589
2360 GTTCCGCTGCGCTGACGTGAGAGTGCCTCAATGAGCCCAAGCTTCTTATCTC 2419
2590 GGATCTGTGGCCATGACAGCAAGAAATTTCCCTGGAACCAATGTTACTGCTCTTA 2649
2420 CAGGCTTCTGAGAGAAACGTCATTTTTCACACAACCTTCCGCTCCCAAGC 2479
2650 CATGTTTCTGATGAGATGTCATTTTGCACATACAGTATATTAAGATTTT 2709
2480 ATCCGAGCAGGAAACCTTACAGCTTCCAGATTTACCCCAAGCAGAGACATATTCG 2539
2710 GAGAGGCTGAGAAAGCATATGATTTACAGATATCTCAGAGAGACACAGCATAGA 2769
2540 TGGCCGAGTGGGCGAGACATATGAGTACGCTGCTGCACTTCTACAGAAATCTC 2599
2770 GTTCTGATCTGGAGAACATATGAACTGCACTTTTGCACACTACCTCAAGAAACCT 2829
2600 TGAGC 2604
2830 GGATC 2834

RESULT 10
BC030688
LOCUS BC030688 3130 bp mRNA linear PRI 30-MAY-2002
DEFINITION Homo sapiens, Similar to dipeptidylpeptidase 8, clone MGC:26191

IMAGE:4822550, mRNA, complete cds.
BC030688
BC030688.1 GI:21265132
KEYWORDS
MGC.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 3130)
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (24-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC project URL: <http://mgc.ncl.nih.gov>
COMMENT
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shliraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Keltman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAP Plate: 33 Row: d Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 18450277.
FEATURES
source
location/Qualifiers
1..3130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:26191 IMAGE:4822550"
/issue_type="Testis"
/clone_1b="NIH_MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
372..2867
/codon_start=1
/product="Similar to dipeptidylpeptidase 8"
/protein_id="AAH30688.1"
/db_xref="GI:21265133"
/translation="MAAMETQQLVEIFETADCEENIESQRPKLEPERYVERYSQ
LKLILADTRKYHGMARAPHDPMFVKRNDPGPSPDRITLYLANSGENENLYESET
PRTINRAVVLMSKPLDLFOATLDYGMYSSEELRERKIGVGLASDYHOGSG
TELFOAGSGIYHVKGSGPOGFTQDLRLPLVETSCPNIRMDERLCPADPDWIAFLHSN
DWISNIVYREERRLTYVHNELANNEEDRSRGVATFVLOEEDRSGVMCPKREMT
PSGKILITVIEENDESEVEITHVSPMLFETRADSPFRPKGTANPVTTFKMSIMT
DAEGRIIDYIKELIOPPEILFEGVEIYIARACMTPEKXANSILDRSOTRLQYVLS
PELFIPEVDVVERQRLIESVPSVPLITYETDIDVINDIHDFHVPQSHSEET
IFASECKGFRHLRYKITSILKESKRSRSGGLPAPSDFCPIKEBIAITSGBEVLGR
HGSNIQVDEVRILYVEFGTKDPLDLBNHLYVYVVEGTATRLTDGYSHSCISQHD
PFSIKYSNOKNHCVSLKSSPEDPCKTKEFNAITLDSAGPLDPDTPPTFSFES
TTGFTLYGMILKYPHDLOGCKKYPYLFITFGGNOVLNNRFGVGFYPLNLTASISGY
VVDINRSGCHRGKLFEGAFKXMAIAGAPVTLWIFDTGTETTERMGPDPONDEGY
LGSVMAQAEKPEPSPENRLLHLGFLDENVAHNTSILSLFLVRAKRPVDLOITPQERH
SIRPESEHDELHLHYLOENLGSRIALAKVI"
BASE COUNT
955 a 626 c 717 g 832 t
ORIGIN
Query Match 24.9%; Score 651.8; DB:9; Length 3130;
Best Local Similarity 58.3%; Pred. No. 2.8e-101;
Matches 1180; Conservative 0; Mismatches 837; Indels 6; Gaps 2;
DB 474 TTTTATGTTGAGCGGATTCCTGAGTCACTTAAAGCTGCTGCCATCCAGAGAAA 533

Kearney, L., Tribouley, C. M., Khan, F. A., Yao, M. G., Baughn, M. R.,
 Azimzal, Y., Elliott, V. S., Nguyen, D. B., Gandhi, A. R., Yang, J.,
 Hernandez, R., Policky, J. L., Lu, D. A., Reddy, R., and Tang, Y. T.,
 Processes
 Patent: WO 0183775-A 16 08-NOV-2001;

Incyte Genomics, Inc. (US)
 Localion/Qualifiers

FEATURES

source
 1. .2510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 376067CB1"
 BASE COUNT 777 a 494 c 527 g 712 t
 ORIGIN

Query Match 21.9%; Score 573.8; DB 6; Length 2510;
 Best Local Similarity 57.4%; Pred. No. 5.4e-88;
 Matches 1138; Conservative 0; Mismatches 822; Indels 21; Gaps 5;

QY 127 CGGACCCGACGAGTCTGGGCTCTGTCACAAAGCGGCCGACGCTTCCAGTTTGT 186
 DB 29 CGATTACAGAAATATCATGGCTACATGATGGCTAAGGCCACCATGATTCTGTTGT 88
 QY 187 GCAGAGAGAGATGATGCTGGGCGCCACGCGCTCTACTACTGGGATGSCATA 246
 DB 89 GAAGAGAAATGATCCAGATGAGCTCTATTCAGACAGAACTATTACCTGCGATGCTGG 148
 QY 247 TGGACCCGACGAGAACCTCCCTCTCTACTCTGATGATTCACAAAGAGTCCGAAAGGC 306
 DB 149 TGAAGACAGAGAAATACACTGTTTATCTGAAATTCACAAACATCATATGAGACGC 208
 QY 307 TCTGCTGCTCTGCTCTGGAAGCAGATGCTGATTCATTTCCAGGCCACGCCACCATG 366
 DB 209 AGCTTAATGCTCTCTGGAAGGCTCTTTGGATCTTTTTCAGGCAACTGAGCATATGG 268
 QY 367 GGTCTACTCGGAGAGAGAGCTGCTGAGGAGCGGAAAGCGCTGGGCTCTCGGCAAT 426
 DB 269 AATGATTTCTCGAAGAAAGAACTATTAAAGAAAGAAACGATTGGACAGTGGAAAT 328
 QY 427 CACCTCTCAGATTCACAGCAGAGAGTGCCCTCTCTCTCTCCAGGCGACAAAGCCT 486
 DB 329 TCGTCTTACGATTAATCCACAGGAAGTGAACATTTCTGTTTCAAGCCGGTATGGAAAT 388
 QY 487 CTTCACCTCTCGGAGCGGCGCAAGAGCTTCAATGCTGCTCCCTATGAAACCGCTGA 546
 DB 389 TTATCTACGTAAGAGATGAGGCGCCACAAAGATTTACGCAACACTTTAAAGGCCCATCT 448
 QY 547 AATCAGAGCCAGTGTCTAGGCGCCGAGTGAACCCCAAAATGSCCTGGCCGACCTGC 606
 DB 449 AGTGAAGACTAGTTGTCACAAACATACGAGATGATCCAAATTAATGSCCTGCTGATCAGA 508
 QY 607 CTCTCTCTCTCATCATATAGACGACCTGTTGGTGGCCAAACATGAGACAGGCGAGA 666
 DB 509 CTGATGTTGCTTTATATACATAGCAAGATATTGGATATCTTAACATGTTAAACGAGAA 568
 QY 667 GCGGCGGCTGACCTTCTGCCACCAAGGTTTATCAATGCTGAGATGACCCCAAGTCTG 726
 DB 569 AAGGAGACTCATATATGTCACAAATGAGCTAGCCAAATGAGAAAGATGACGATCAGC 628
 QY 727 GGTGTGGGCGCTGTCATACAGGAAGATTCGACCCCTTACAGGGGATGAGTGGTGG 786
 DB 629 TGGAGTGGTACCTTTGTTCTCCAAAGAAATTTGATTAATTTCTGGGCTATTTGGTGG 688
 QY 787 CCCACAGGCTCTGAGGAGTTTCAGAGGCGCTCAAGAGCTGCGAATCTGATAGAGA 846
 DB 689 TCCAAAG--CTGAAGAACTCCAGTGGTGGTAAATTTCTTGAATCTTATGAAGA 745
 QY 847 AGTCATAGTCCGAGGTGGAGTCAATTCAGTCCCTCTCTCGGCTGAGAAAGG-A 905
 DB 746 AATGATGATCTGAGGTGAATTTATCATGTTTACATCCCTATGTTGAACAAGCA 805
 QY 906 AAGAGGAGCTGATGCGTCCCGAGGAGGAGCAAGAAATCCCAAGATTTGCTTGAAG 965
 DB 806 GGGCAGATTCATTCGTTATCTTAAGAGGTACAGCAAAATCCTAAGTCACTTTAAGA 865

QY 966 TGGCTGATTTCCAGACTGACAGCCAGGCGAAGATCTCTGACCCAGAGAAAGAGCTGG 1025
 DB 866 TGTACAGAAATATGATTTGATGTCGAAGGAAAGATCATATGATCATATGAAAGAACTAA 925
 QY 1026 TGCAGCCCTTCAGCTGCTGTTTCCGAGGTGAGTACATGCCAGGCGCGGTGAGCC 1085
 DB 926 TTCAACCTTTTGAATCTTATTTGAAGAGATTGAATATATTTCCAGAGCTGATGAGCTC 985
 QY 1086 GGGATGGCAATATGCTGGGCGATGCTCTGACCGCGGCCAGAGCTGGCTCCACTGCG 1145
 DB 986 CTGAGGAAATATGCTGGTGGCTCTACTAGATGCTCCGAGCTGCGCTGCAATAG 1045
 QY 1146 TCTCTCTCCCGCCGCGCTCTGTTCCAGCAGAGATGAGAGAGCGCTAGCCT 1205
 DB 1046 TGTGATATCAGCTGATATTTATTTCCAGTACAGATGATGTTATGGAAGGACAGAGC 1105
 QY 1206 CTGCGAGCTGTCCCGAGGATGTCAGCCGATGCTGTGATGAGAGGATCAGACAG 1265
 DB 1106 TCATTTGAGTACAGCTGATGATTTCTGACGCGCACTAATTTATGGAAGAAACAGACA 1165
 QY 1266 TCTGATGCAATGTTCAATGATCTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGC 1325
 DB 1166 TCTGATTAATATTCATGATGATCTTTCATGTTTCCCAAGCTCACAGAGAG--GAAA 1222
 QY 1326 TCTGCTTTCTCCGCGCAATGATGCAAGACCGCTTCTGCTTGTATGCAAAAGTACCG 1385
 DB 1223 TTGAGTTTATTTTGGCTCTGATGCAAAACAGGTTTCCGTCATTTATCAAAATTTACAT 1282
 QY 1386 CCGTTTTAAATCCAGGCTACAGATTTGAGTGAAGCCCTTCAGCCCCGGGGAAGATGAT 1445
 DB 1283 CTATTTTAA-----GGAAGCAAAATATAAGATTCAGTGGTGGGCTGCTGCT 1332
 QY 1446 TTGATGCCCCATTAGAAGATGCTCTGACAGCGGTGAATGGAGGTTTGGCA 1505
 DB 1333 CCACGTCATCTGATGATGATCATTCATGAGATCTGAGAACTCATCTGATATGTG 1392
 QY 1506 GGCAGGCTCCA---AGATCTGGGTCAATGAGAGAACCAAGCTGGTGTACTTCCAGGGC 1561
 DB 1393 GTGACACATATATGATGATTCAGAGTTGATGAGTGAAGAGCTGGTATATTTTGAAGGC 1452
 QY 1562 ACCAGAGACGCGCTGAGACACACTCTGCTGATGCTAGATGAGAGCGCGCGAG 1621
 DB 1453 ACCAAAGACTCCCTTTAGAGCATCTGATGATGATGATGATGATGATGATGATGATG 1512
 QY 1622 ATGCTAGGCTACACAGCGCGCTCTCCATGATGCTCCATGAGGCGCAAGACTGAGC 1681
 DB 1513 GTGACAGGCTGAGTACCGTGGCTACTACATCTTCTGCTGCTGCTGCTGCTGCTGCTG 1572
 QY 1682 ATGTCGTCAGGCTACAGACAGCTGAGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1741
 DB 1573 TTCTTTATTAAGTATGATATGTAACAGAAAGATTCACACATGCTGCTGCTGCTGCTGCTG 1632
 QY 1742 AGGCGCCCGAGAGACAGCCCTGCAAGAGAGCGCGCTTCTGCGCTGCTGCTGCTGCTG 1801
 DB 1633 TCAAGTCTGAGAGATGAGCCCACTTCAAAACAAAGAAATTTTGGGCGCCACATTTGGAT 1692
 QY 1802 GAGAGAGCTGCGCGCGGATTAATGTTCCAGAGATGCTCCATTTCCACAGGCGCTG 1861
 DB 1693 TACGAGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1752
 QY 1862 GATGTGCGGCTCTACGAGATGATCTCAAGAGCCGCTGCTGCTGCTGCTGCTGCTGCTG 1921
 DB 1753 GGAATTACATTTGATGAGATGCTCTCAAGAGCTCAATGATCTACAGCTGGAAGAAATAT 1812
 QY 1922 CCACCGCTCTCTTTGATGAGAGGCGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1981
 DB 1813 CCTACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1872
 QY 1982 GGCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2041
 DB 1873 GAGTCAATATTTTCCGCTTGAATTACCTTACCTCTCTGATGCTGCTGCTGCTGCTGCTG 1932

[illegible]

Db	576	TCCTGGAAAGCCTCTTTTGGATCTTTTTCAGGCAACACTGAGACTATGGAATGATATTCTTCGA	635
Qy	380	GAGGAGAGAGCTGCTGAGGAGGAGCGGAAACGCGCTGGGGGCTCTTCGCGATCACTCCTCAAGAC	439
Db	636	GAAGAGAACATATTAAAGAGAAAGAAAACGCAATGGAAAGCTGGAAATTTGCTTTACGAT	695
Qy	440	TTCCACAGCGAGAGTGGCCCTTCCTTCCTTCACAGCCAGCAACAGCCTCTTCCACTGTGCC	499
Db	696	TATCAACCAAGGAATGGAAACATTTCCTGTTTCAAGCCGGTAGTGGAATTTATCAAGTAA	755
Qy	500	GAGGGGGCAAGAAACGGCTTCATGCTGTCGCCATGAAACCGCTGGAATCAAGCCGAG	559
Db	756	GATGAGAGGCCACAGAGATTTACGCAACAACCTTTTAGGCCCAATCTAGTGAANAATGCT	815
Qy	560	TGCTCAGAGGCCCGGATGAGACCCCAAAATCTGCCCTGGCGACCCCTGCTTCTTCCTTC	619
Db	816	TGTGCCAACAATAGGAGATGGATTCCAAAATTTATGCCCTGCTGATCCAGACTGGATTGCTTTT	875
Qy	620	ATCAATATACAGCGACTGTGGGTGGCCAAACATCGAGACAGCGGAGAGCGCGGCTGAC	679
Db	876	ATACATATGACACGATTTATTTGATATCTAACATTCGTACCCAGGAGAAAGAGAGACTACT	935
Qy	680	TTTCGCCAACAAAGTTTATCCATATGTCGTGAGTANCCCAACTCGCGGGTGTGGCAAC	739
Db	936	TATGTGACACATGTGCTAGGCAACATGAGAAAGATGCCAGATCAGCTGGAGTGTGCTACC	995
Qy	740	TTTCGTCTAACAGAAAGATTGCGACCGGCTTCACGTGGGTACTGTGGTGGCCCAACGCTCC	799
Db	996	TTTGTCTCCAGAAAGATTTGATATATATCTGTGGGTATGTGGTGTGTCCAAAAG---CT	1052
Qy	800	TGGGAAGTTTCAAGAGGCGCTCAAGACGCTGCGAATCTGTATGAGAAAGTATGATGCTC	859
Db	1053	GAACAACACTCCCACTGGTGGTAAATTTCTTGAATTTCTATGAGAAATGATGTAATCT	1112
Qy	880	GAGTGTGAGGTCAATCAACCTCCCTCTCTCGCGCTTAGAAGAAAGAAACGAGACTCGAT	919
Db	1113	GAGGTGGAATTAATCAATGTTAATCTCCCTATGTTGGAAACAGGAGGCGAGATTCATTC	1172
Qy	920	CGGTACCCCAAGAGAGACACAGAAATCCCAAGATGCTTAAACGGCTGAGTTCAG	979
Db	1173	CGTTATCTTAACAGAGGTATACAGCAAAATCTTAAGTACTTTTAACATGTCAAAATATATG	1232
Qy	980	ACTGACAGCCAGGCGAGATCGTCTGAGCCACAGAGAAAGAGACTGTGACGCCCTCAGC	1039
Db	1233	ATTGATGCTGAAAGAGATCATATGATATGTCATATGATAGAACTAATTCACACTTTTGA	1292
Qy	1040	TGCGTGTCCCGAAGGTGGAGTACATCGCCAGGCGCGGCTGGACCCGGGATGGCAATATC	1099
Db	1293	ATTCTATTTGAAAGAGTTGATATATTGGCAGAGCTGATGACATCTGAGGGAATAAT	1352
Qy	1100	GCGTGGGCGATGTTCCGTGGAGCGGGCCCGCAGAGTGGCTCGACCTGCTCTCTCCCGCG	1159
Db	1353	GCTTGTGTCACTACTACTAGTATGCTGCCAGACTCGCTTCAACATATGTGTATCTCACCT	1412
Qy	1160	GCCCTGTTTCATCCGACACAGAGAATGAGAGACAGCGGCTGACCTCTGCCAGAGCTGT	1219
Db	1413	GAATTAATTTATCCAGTAGAAGAAATGATGTTATGGAAGGCAAGACTAATGAGTCAATG	1472
Qy	1220	CCCAAGATGTCCAGCGGTATGTGGTGTACAGAGAGGTACCAACGCTGTGATCAATGTT	1279
Db	1473	CCTGATCTGTGAGCGCACTAATTAATCTATGAGAAACAAACAGACATCTGGTAAATATC	1532
Qy	1280	CATGCAATCTTCTATCCCTTCCCAATCGAGAGGAGAGAGAGACTGCTGCTTCTCCG	1339
Db	1533	CATGACATCTTTCTATGTTTTTCCCAAGCTCAGCAAGAG---GAATTTGAGTTATTTTT	1589
Qy	1340	GCCAAATGATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCACCGCGTTTTAAATCC	1399
Db	1590	GCGTCTGTATGCAAAACAGAGTTTCCGTCATTTATACAAATTAATCATTTATTTAAAGAA	1649
Qy	1400	CAGGCTACGATTTGAGTGTAGCCCTTCAGCCCGGGGGAAGATGAATTTAAGTCCCACTT	1459

Db 1650 AGCAAAATATAAAGATCCAGTGGTGGGCTGCTGCTCCAAAGTGAATTTCAAGTCTATC 1709
QY 1460 AAGGAAGAGATGCTCTGACACGCGGTGAATGGAGAGTTTGGCGAGCAGCGCTCCAG 1519
Db 1710 AAGAGAGATATACCAATTTACAGTGGTGAATGGAAAGTCTTGGCCGCGCATGATCTAAT 1769
QY 1520 ATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGGCACCAGACAGACGCGCTG 1579
Db 1770 ATCCAAGTTGATGAAGTCAAGAGCTGGTATATTTTGAAGCACCAGCAAGACTCCCTTTA 1829
QY 1580 GAGCACCACCTTACGAGTGCAGCTATGAGCGCGCCGCGCAGATCGTACGCTCACCAG 1639
Db 1830 GACATCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1889
QY 1640 CCGGCTGCTCCATAGCTCTCCATGAGCCAGAACTTGCATGTTCTGACGACTTAC 1699
Db 1890 CGTGGCTACATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1949
QY 1700 AGCAGCTGAGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1759
Db 1950 AGTAAACAGAGATCCACACTGTGTGCTGCTTACAGCTATCAAGTCTCTGAAGATGAC 2009
QY 1760 CCGCTGACAGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1819
Db 2010 CCACTGTTGAAAAGAGATTTTGGCCACCATTTTGGATTCAGCAGGCTCTCTTCT 2069
QY 1820 GATTATGCTCTCCAGATCTTCCATTTCCACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1879
Db 2070 GACTATATCTCTCCAGAAATTTTCTCTTTGAAAGTACTGCTGATTTACATTTGATGGG 2129
QY 1880 ATGATCTACAGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1939
Db 2130 ATGCTACAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2189
QY 1940 TATGAGAGCCCCCAGGT 1956
Db 2190 TATGTGCTCTCAGGT 2206

RESULT 13
AK000290
LOCUS AK000290 2778 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens CDNA FLJ20283 f1s, clone HEP04088.
VERSION AK000290.1 GI:7020272
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:HePG2 CDNA to mRNA, clone_l1b:HEP
clone:HEP04088.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,
Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2778)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,
Minato-ku Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
source Location/Qualifiers
1..2778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP04088"
/cell_line="HepG2"
/cell_type="hepatoma"
/clone_l1b="HEP"
/note="cloning vector pME18SFL3"
163..1572
/note="unnamed protein product"
/codon_start=1
/protein_id="BA91059.1"
/db_xref="GI:7020273"
/translation="MAAMETBOLGVEIFETADCEENIESODRPRKLEPYVERYSMSO
LKILADPTKHYGHMAKAPHDMPYKRPNDPPGSPDRXYLLAMSGEENELFYSEI
PKTINAAVLMISMRFLDLPDAILDYGATVREBELREKRRTGYGISTYHQSG
TFLPQSGSIVHKDGGPOGFTQOPLRPNLVRESCPNIRMDPLKCPADPDWIAFIHSN
DIWISNIVREERRLTYVHNEELANMEDARSAGVAVLOEEDRYSQWMCPRKATTT
PSGKILRLIYEENDESEVEIHTVSPLMETRADSFRYKGTANPKVTFKMSIEMI
DAEGRIDVIDKELIQPEILFEGVEYIARAGWTPGKAYMSIILDRSOTRIQVILIS
PELFIPEEDVMEBRORLIESVPSVPLIYEETDIWINDIHIVFPQSHHEEIEF
IRASEKTEFRHLIKITST"

BASE COUNT 856 a 562 c 615 g 745 t
ORIGIN

Query Match 21.3%; Score 557.8; DB 9; Length 2778;
Best Local Similarity 57.2%; Pred. No. 2.8e-85;
Matches 1073; Conservative 0; Mismatches 797; Indels 7; Gaps 3;

QY 80 TTCCAGGTGAGAGAGCACTGCTGGAGCGGCTCCGAGCATATCCAGCGAGCCGCAAG 139
Db 265 TTTTATGTTGAGGGGTATCTCTGAGTACAGCTTAAAAAGCTGCTGCCATACAGAAAA 324
QY 140 TACTCGGCTCATTTGTCAACAAGCGCCCGCAGCAGCTTCCAGTTTGTCAAGAAAGGAT 139
Db 325 TATCATGCGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
QY 200 GAGCTGGGCGCCACCTCCAGCCGCTCTACTACTGATGATGATGATGATGATGATGATGAT 289
Db 385 CCAATGAGACTCTATTCAGACAGAAATCTATTTACTTCCATGCTGTGTGAGAACAGAGAA 444
QY 260 AACTCCCTCTACTCTGAGATTTCCAGAGAGCTCCGGAAGAGCGCTGCTGCTCTG 319
Db 445 AATACACTGTTTATTTCTGAATTTCCCAAAATATCAATGAGACAGCAGCTTAATGCTC 504
QY 320 TCTGTGAGAGAGATGCTGATTCATTTCCAGGCCAGCGCCACCATGAGGCTACTCTCGG 379
Db 505 TCTTGGAGGCTCTTTTGGATCTTTTCAAGCAACACTGGACTATGATGATGATGATGATGAT 564
QY 380 GAGGAGAGCTGCTGAGAGGAGGAGAAAGCGCTGGGGCTTTCGAGCATCACTCTACGAC 439
Db 565 GAAGAAGAACTATTAGAGAGAAAGAAAGCAATTCGGAACATGCGAATGCTTCTTACGAT 624
QY 440 TTCCACAGCAGAGTGGCTCTCTCTCTCCAGGCCAGCAACAGCTTTCACACTGTCG 499
Db 625 TATCACCAAGAGAGTGAACATTTCTGTTCAAGCGGCTAGTGAATTTATCACGTAATAA 684
QY 500 GAGCGCGGCAAGAACGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 559
Db 685 GATGAGAGGCGCACAGAGATTTTACGCAACAACCTTTTAAAGGCCCAATCTAGTGAACATGAT 744
QY 560 TGCTCAGGGCGCGGATGAGAGCCCAAAATGTGCTGCGGAGCCGCTGCTCTCTCTCTG 619
Db 745 TGTCCCAACATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 804
QY 620 ATCAATAACAGCACTGTGGTGGCCAAATGAGACAGCAGAGAGAGCGCGCTGAC 679
Db 805 ATACATAGCAAGATATTTGGATATGATCAATGCTGAACAGAGAAGAGAGACTACT 864
QY 680 TTTCGCCACCAAGTTATTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 739
Db 865 TATGTGCACAATGAGCTAGCAACATGAGAAAGAGATGCCAGTACGTGAGTGGCTTACC 924

IFASECKTGRHLKYITSLIKESKYNRSSGGLPAPSDFKPIKEEIAITISGEWEVLGR
HGSNIQVDEVRRLVYFEGTDSPLLEHLHYVSYVNGEVRLLDRGSHSCJISOHD
FPIKSNOKNPHVCISLYKLSPEDEPTCKTKEFMAITLDSAPLDYRPPPELSESES
TTGFTLXGMLYKHPDLOPKKYPVLYFYGGPVALIAGATVIMIPDYTYTRYKMH
PDNEDGTILGSAVMOAERKPEPSENRLLHGLDEHLEHVAHSTILSFVLRAGKPPD
LQIYPERHSIRVPESEHEHLLHYLHLENLDSRIAALVKI*

BASE COUNT 796 a 564 c 592 g 716 t
ORIGIN

Query Match 16.6%; Score 434.6; DB 6; Length 2668;
Best Local Similarity 58.1%; Pred. No. 2,7e-64;
Matches 805; Conservative 0; Mismatches 574; Indels 6; Gaps 2;

Qy 572 CGAGTGGACCCCAAAATCTGCGCCGACCCCTTCTCTCTCATCAATAACAGC 631
Db 654 CGAGTGGATCCAAATTTATGCTGCTGATCCAGACTGATGCTTTTATACATAGCAAC 713
Qy 632 GACCTGTGGTGGCCCAATCGAGACGAGGCGCGCTGACCTTCTGCCACCAA 691
Db 714 GATATTGGATATCTAAGATGTAACAGAGAGAAAGAGACATCATTTATGTGCACAT 773
Qy 692 GGTATTACCAATGTCCTGGATGACCCCAAGTCTGCGGTGGCCACCTTCTCATACAG 751
Db 774 GAGCTACCCCAACATGAGAAAGATGCCAGATCAGCTGAGTCCCTACTTTGTCTCCAA 833
Qy 752 GAAGAGTTGCGCCGCTCAGTGGTACGTCGTGCGCCCAAGCCCTCTGGGAAGTTCA 811
Db 834 GAAGAAATTTGATATATTTCTGGCTATTTGGTGTCCAAAG--CTGAACAACATCC 890
Qy 812 GAGGCGCTCAAGACGCTGCGAATCTGTATAGAGAAAGTCATGTAGTCCGAGTGGAGTTC 871
Db 891 AGTGTGGTAAATTTCTTGAATCTATATGAAAGAAATGATGATCTGAGAGTGGAAAT 950
Qy 872 ATTACGTCCTCTCTCTGCGCTAGAGAAAGAAAGACGAGCTGTATCGGTACCCAGG 931
Db 951 ATTATGTATACATCCCTATTTGTAACAAAGAGGAGGAGATTCCTTATCTTAA 1010
Qy 932 ACAGGCGCAAGATCCCAAGATGSCCTTGAATGCTGTGATCCAGACAGACAGCAG 991
Db 1011 ACAGGTACAGCAAAATCTTAAAGTCACTTTTAAATGTGCAAAATTAATGATGCTGAA 1070
Qy 992 GGCAGATTCGCTCTGACCCAGAGAGAGAGAGTGTGACGCCCTTACGCTGCTTCCCG 1051
Db 1071 GGAAGATTCATAGATGTCATGATAGAGAACTATTCACCTTTTGAATTTCTATTTGAA 1130
Qy 1052 AAGGTGAGTACATGCGCAGGCGCGGTGACCCGGAATGCAATACGCTGGGCCATG 1111
Db 1131 GGAATTAATATATTTGCCAGAGCTGATGACTCCTGAGGAAATATGTTGGTCCATC 1190
Qy 1112 TTCTGTGACCGGCGCCAGCAGTGGCTCAGCTCTCTCTCCCGCGGCTGTTCATC 1171
Db 1191 CTACTAATGCTGCTCCAGACTGCTTACAGTATGTTGATCTCAGCAATTTATTTATC 1250
Qy 1172 CCGAGCACAGAGATGAGAGACAGCGCTAGCTCTGTGCCAGAGCTGTCCCAAGAAATGTC 1231
Db 1251 CCACTAGAAAGTATGATTTATGAGAAAGAGAGACTCATGTGATGCTGATTTCTGTG 1310
Qy 1232 CAGCCGATGTGTGTACGAGAGAGAGTCAACAAGCTGTGATCAATGTTTATGACATCTTC 1291
Db 1311 ACGCCACTAATTTATCTATGAGAAACAACAGACATCTGATTAATATCATGACATCTTT 1370
Qy 1292 TATCCCTTCCCGCAATCAGAGAGAGAGAGAGCTGTCTTCTCCGCGCAATGAATGC 1351
Db 1371 CATGTTTTTCCCAAGAGTCAAGAGAG--GAATTTGAGTTTATTTTGGCTGTGAATGC 1427
Qy 1352 AAGACCGGCTTCTGCAATTTGTACAAAGTCAACCCGCTTTTAAATCCAGGGCTACGAT 1411
Db 1428 AAACAGAGTTTCCGTCATTTTACAAATTTACATCTATTTTAAAGAAAGCAATATATAA 1487
Qy 1412 TGAAGTGAAGCCCTTCAAGCCCGGGAAGATGAATTTAAAGTCCCATTTAAGGAAGATTT 1471
Db 1488 CGATTCAGTGGTGGGCTGCTCCTCAAGTATTTCAAGTCTCTATCAAGAGAGAGATA 1547

Qy 1472 GCTGTGACACGCGTGAATGGAGAGTTTGGCGAGGACAGGCTCCAGATCTGGGTCAAT 1531
Db 1548 GCAATTTACAGATGGTGAATGGAGAGTTCTTGGCCGATGATCATTAATTCAGAGTTGAT 1607
Qy 1532 GAGAGACCAAGCTGTGTACTTCCAGGGCACCAAGAGACGCGCTGGAGACACCACTC 1591
Db 1608 GAACTCAGAAAGCTGTATATTTTGAAGGCCACCAAGACTCCCTTTAAGACATACCTG 1667
Qy 1592 TACGTGTACACTATGAGAGCGCGCGGAGATCTGAGCCCTACACAGCCCGGCTTCTCC 1651
Db 1668 TACGTATGACATTTAGTAATCTGAGAGAGTGAACAAAGCTGACTGACCCGTGACTCA 1727
Qy 1652 CATAGCTGCTCCATGAGGACAGACACTGCAATGTTGTCACGCCACTAGAGAGGTGAGC 1711
Db 1728 CATCTGTGCTCATCAGTCACACTGTGACTTTTATTAAGTATAGTATACAGAGAG 1787
Qy 1712 ACGCGCCCTGCGTGCACAGTCTACAGCTGAGCGCCCGACGAGACCCCTGCACAAAG 1771
Db 1788 AATCCACAGCTGTGTGCTCCTTTACAGATATCAAGTCTGAAGATGACCCAACTTGCAA 1847
Qy 1772 CAGCCCGCTTCTGTGGCTAGCATGTGAGAGCAGCCAGCTCCCCCGGATTAATGTTCT 1831
Db 1848 ACAAAGGAATTTTGGCCACCATTTTGGATTACAGCAGGTCTCTTCTGACTATATCTCT 1907
Qy 1832 CCAGAGATCTTCCATTTCCACACAGCGGCTCGGATGTGGCGCTCTAAGCATGATCTACAAG 1891
Db 1908 CCAGAAATTTCTCTTTTGAAGTACTACTGATTTACATTTGATGAGATCTCTACAG 1967
Qy 1892 CCCCAGCCTTGCACAGGAGAGAGACCCACCGCTCTCTTGTATATAGAGAGGCC 1951
Db 1968 CCTATGATCTACAGCTCGGAAGAAATATCTTACTGTGCTGTCTATATGATGTTCT 2027
Qy 1952 CAGGT 1956
Db 2028 CAGGT 2032

Search completed: December 12, 2002, 09:56:58
Job time : 4513 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 07:16:06 : Search time 360 Seconds

(without alignments)
16370.777 Million cell updates/sec

Title: US-09-976-674-4

Perfect score: 2617
Sequence: 1 caagctaccatgcgcacaca.....ctcgagcgccgcgcatccg 2617

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq-101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2617	100.0	2617	24	ABK83323	cdna encoding huma
2	2592.4	99.1	24	ABK83335	cdna encoding huma	
3	2592.4	99.1	4302	24	ABK83333	cdna encoding huma
4	2587.6	98.9	3024	24	ABD38954	Human dipeptidyl p
5	2504.4	95.7	4180	24	ABK83339	cdna encoding huma
6	2504.4	95.7	4263	24	ABK83338	cdna encoding huma
7	2488.6	95.1	2495	24	ABD38957	Human dipeptidyl p
8	2296.4	87.7	4076	24	ABK83337	cdna encoding huma
9	2296.4	87.7	4159	24	ABK83336	cdna encoding huma

10	2288.2	87.4	2801	22	AA157896	Human polynucleoti
11	2208.4	84.4	4037	24	ABK83341	CDNA encoding huma
12	2208.4	84.4	4120	24	ABK83340	CDNA encoding huma
13	2091.4	79.9	3262	22	AA157880	Human polynucleoti
14	1936.4	74.0	3287	24	AA157880	Human polynucleoti
15	1869.6	71.4	2461	21	AACT5835	Alternative versio
16	1867.4	71.4	2751	24	AACT5835	Human ORF ORF1390
17	1850.4	70.7	2982	22	AA156666	Murine dipeptidyl
18	1585.6	60.6	3047	24	ABK69113	Human polynucleoti
19	1584.4	60.5	2952	24	ABK69113	DNA encoding huma
20	1429.8	54.6	2411	24	ABK69090	DNA encoding huma
21	1426.6	54.5	2079	21	AAA37672	Human peptidase, H
22	1386.2	53.0	2034	22	AA159682	Human polynucleoti
23	1316	50.3	2027	24	ABK69114	DNA encoding huma
24	1308	50.0	2027	21	AACT7137	Human ORF ORF2632
25	862.6	33.0	2643	24	AAH99935	Coding sequence of
26	862.6	33.0	2671	24	ABK12892	CDNA encoding huma
27	862.6	33.0	3106	24	ABK12892	Human protease PRT
28	862.6	33.0	3143	24	AAH99934	CDNA encoding 2195
29	861	32.9	3120	22	AAH85694	Nucleotide sequenc
30	861	32.9	3120	24	AAH85694	Human dipeptidyl P
31	850.6	32.5	4829	24	ABK83327	CDNA encoding huma
32	658.2	25.2	4685	24	ABK83332	CDNA encoding huma
33	614.6	23.5	4676	24	ABK83331	CDNA encoding huma
34	573.8	21.9	2510	24	AAH23843	Human protease PRT
35	568.2	21.7	2842	24	ABK59774	Novel human coding
36	566.4	21.6	4523	24	ABK83325	CDNA encoding huma
37	563	21.5	1048	22	AAH41004	CDNA encoding nove
38	518	19.8	662	22	AAK92083	Human CDNA clone r
39	518	19.8	662	22	AAK93366	Human full-length
40	518	19.8	1748	22	AAH94819	Human CDNA sequenc
41	516.6	19.7	2161	22	AAH15009	Novel human coding
42	445.6	17.0	4309	24	ABK83328	CDNA encoding huma
43	430.4	16.6	2668	24	ABK59775	Novel human coding
44	394.6	14.9	3713	23	ABH10425	Drosophila melanog
45	390.4	14.9	3783	23	ABH06641	Drosophila melanog

ALIGNMENTS

RESULT 1	ABK83323	standard; CDNA; 2617 BP.
XX	ABK83323:	
XX	12-AUG-2002 (first entry)	
XX	CDNA encoding human DPPIV related serine protease DPP-2.	
XX	Human; serine protease; dipeptidyl peptidase IV-related protein; DPP;	
XX	DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;	
XX	diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;	
XX	heart failure; hypertension; urinary retention; osteoporosis; cancer;	
XX	ulcer; allergy; cancer; psychotic disorder; neurological disorder;	
XX	dyskinesia; reproductive disorder; inflammatory disorder;	
XX	metabolic disorder; gene; ss.	
XX	Homo sapiens.	
XX	WO200231134-A2.	
XX	18-APR-2002.	
XX	12-OCT-2001; 2001WO-US31874.	
XX	12-OCT-2000; 2000US-240117P.	
XX	(FERR) FERRING BV.	
XX	Ol S, Akinsanya KO, Riviere PJ, Juntlen J;	

DR MPI: 2002-444178/47.
XX P-PSDB: AB651592.
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX -
XX Claim 1: Page 56-57; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPPR proteins.
XX
XX Sequence 2617 BP; 558 A; 830 C; 736 G; 493 T; 0 other:
SQ
Query Match 100.0%; Score 2617; DB 24; Length 2617;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 CAAGCTTACCATTGGCCACACCGGGACCCCAACGCGCGACCGAGCGACCGCCAC 60
DB 1 CAAGCTTACCATTGGCCACACCGGGACCCCAACGCGCGACCGAGCGACCGCCAC 60
OY 61 AGATGACCGCGCGCGCGCGCTCCAGGTGCAAGAACACTGCTGGAGCGGCTCCGGACAT 120
DB 61 AGATGACCGCGCGCGCGCGCTCCAGGTGCAAGAACACTGCTGGAGCGGCTCCGGACAT 120
OY 121 CATCCACGCGACCGCAAGTACTCTGCGGCTCATATTGCAACAGCGCCCGCACGACTTCCA 180
DB 121 CATCCACGCGACCGCAAGTACTCTGCGGCTCATATTGCAACAGCGCCCGCACGACTTCCA 180
OY 181 GTTTGTGCAAGAACGAGTGTGCGGCCCACTCCACCGCTTACTACTCTGGGAAT 240
DB 181 GTTTGTGCAAGAACGAGTGTGCGGCCCACTCCACCGCTTACTACTCTGGGAAT 240
OY 241 GGCATTGCGACGCGAGAACCTCCTCTACTCTGAGATTCCCAAGAAAGTCCGGAA 300
DB 241 GGCATTGCGACGCGAGAACCTCCTCTACTCTGAGATTCCCAAGAAAGTCCGGAA 300
OY 301 AGAGGCTGTGCTCTCTCTGTAAGACAGATGTGATCATTTCCAGGCCACGCCCA 360
DB 301 AGAGGCTGTGCTCTCTCTGTAAGACAGATGTGATCATTTCCAGGCCACGCCCA 360
OY 361 CCATGGGCTTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAAGCCCTGGGGTCTT 420
DB 361 CCATGGGCTTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAAGCCCTGGGGTCTT 420
OY 421 CGGCATCACCCTCTAGATTTCCACAGCAGAGTGGCTTCTTCCAGGCCAGCAA 480
DB 421 CGGCATCACCCTCTAGATTTCCACAGCAGAGTGGCTTCTTCCAGGCCAGCAA 480
OY 481 CAGGCTTCTCACTGTGCGAGCGCGGCAAGAAAGGCTTCATGGTTCCCTTGAAC 540
DB 481 CAGGCTTCTCACTGTGCGAGCGCGGCAAGAAAGGCTTCATGGTTCCCTTGAAC 540
OY 541 GCTGGAATCAAGACCAAGTGTGAGGCGCCGGATGAGACCCCAAAATCTGCCCTGCCA 600
DB 541 GCTGGAATCAAGACCAAGTGTGAGGCGCCGGATGAGACCCCAAAATCTGCCCTGCCA 600
OY 601 CCTGCTTCTCTCTTATCATTAACAGGACCTGTGGGTGGCCAAATGAGACAG 660
DB 601 CCTGCTTCTCTCTTATCATTAACAGGACCTGTGGGTGGCCAAATGAGACAG 660

DB 601 CCTGCTTCTCTCTTATCATTAACAGGACCTGTGGGTGGCCAAATGAGACAG 660
OY 661 CGAGAGCGCGGCTGACCTTGTGCCACCAAGGTTTATCCAAATGCTCGATGACCCAA 720
DB 661 CGAGAGCGCGGCTGACCTTGTGCCACCAAGGTTTATCCAAATGCTCGATGACCCAA 720
OY 721 GTCTGGGGTGTGGCCACTTGTCTATACAGAAAGTTGACCGCTTCACTGGTACTG 780
DB 721 GTCTGGGGTGTGGCCACTTGTCTATACAGAAAGTTGACCGCTTCACTGGTACTG 780
OY 781 GTGTGCCCCACAGCCCTCTGGGAAGTTTCAGAGGCGCTCAACAGCTCGAATCTGTA 840
DB 781 GTGTGCCCCACAGCCCTCTGGGAAGTTTCAGAGGCGCTCAACAGCTCGAATCTGTA 840
OY 841 TGAGAGTGTGATGATCCGAGGTGAGGTGATTCAGTCCCTCTCTGCGCTAGAA 900
DB 841 TGAGAGTGTGATGATCCGAGGTGAGGTGATTCAGTCCCTCTCTGCGCTAGAA 900
OY 901 AAGGAAGCGAGCTGTATCGGTACCCAGAGACGAGCAGCAAGATCCCAAGATTGCTT 960
DB 901 AAGGAAGCGAGCTGTATCGGTACCCAGAGACGAGCAGCAAGATCCCAAGATTGCTT 960
OY 961 GAACCTGGGTGATGATCCAGACTGACAGCCAGGCGAGATGCTCGACCCAGGAGAA 1020
DB 961 GAACCTGGGTGATGATCCAGACTGACAGCCAGGCGAGATGCTCGACCCAGGAGAA 1020
OY 1021 GCTGTGACAGCCCTTACAGCTGCTGTTCGCCGAAGGTGAGTATACATGCGCAGGCGG 1080
DB 1021 GCTGTGACAGCCCTTACAGCTGCTGTTCGCCGAAGGTGAGTATACATGCGCAGGCGG 1080
OY 1081 GACCGGGATGCGCAATACGCTGGGCAATGTTCTGTGACCGGCGCCAGCAGTGGCTCCA 1140
DB 1081 GACCGGGATGCGCAATACGCTGGGCAATGTTCTGTGACCGGCGCCAGCAGTGGCTCCA 1140
OY 1141 GCTGTGCTCTCTCCCGCGGCGCTGTTCATCCGAGACACAGAAATGAGAGAGCGGCT 1200
DB 1141 GCTGTGCTCTCTCCCGCGGCGCTGTTCATCCGAGACACAGAAATGAGAGAGCGGCT 1200
OY 1201 AGCTCTGCGCAGAGCTGTGCTCCAGAAATGTCAGCCGTATGTGTTAGAGAGGTAC 1260
DB 1201 AGCTCTGCGCAGAGCTGTGCTCCAGAAATGTCAGCCGTATGTGTTAGAGAGGTAC 1260
OY 1261 CAACGTCTGATCATATGTTATGACATCTTCTATCCCTTCCCAATACAGAGGAGAGA 1320
DB 1261 CAACGTCTGATCATATGTTATGACATCTTCTATCCCTTCCCAATACAGAGGAGAGA 1320
OY 1321 CGAGCTCTGCTTCTCCGCGGCGCAATGAATGCAAGACCGGCTGTGCAATTTGACAAG 1380
DB 1321 CGAGCTCTGCTTCTCCGCGGCGCAATGAATGCAAGACCGGCTGTGCAATTTGACAAG 1380
OY 1381 CACCGCGCTTTTAAATCCAGAGGCTACGATTGGAATGAGCCCTTACGCCCGGGAGA 1440
DB 1381 CACCGCGCTTTTAAATCCAGAGGCTACGATTGGAATGAGCCCTTACGCCCGGGAGA 1440
OY 1441 TGAATTTAAGTCCCATTTAAGAGAGATGCTGTGACCAAGCGTGAATGGAGTTT 1500
DB 1441 TGAATTTAAGTCCCATTTAAGAGAGATGCTGTGACCAAGCGTGAATGGAGTTT 1500
OY 1501 GGGAGAGCGGCTCAATCGGTGCAATGAGAGACCAAGCTGTACTTCCAGGG 1560
DB 1501 GGGAGAGCGGCTCAATCGGTGCAATGAGAGACCAAGCTGTACTTCCAGGG 1560
OY 1561 CACCAAGGACAGCGGCTGAGACACCTTACAGTGTACAGTATGAGGCGCGGCA 1620
DB 1561 CACCAAGGACAGCGGCTGAGACACCTTACAGTGTACAGTATGAGGCGCGGCA 1620
OY 1621 GATGTAAGCTTACAGCGCGGCTTTCATATGCTCTCCATGAGCCAGAACTTGA 1680
DB 1621 GATGTAAGCTTACAGCGCGGCTTTCATATGCTCTCCATGAGCCAGAACTTGA 1680
OY 1681 CATGTTGCTACAGCACTAAGACAGCGTGTGAGCAAGCGGCGCTGCTACAGAGCT 1740
DB 1681 CATGTTGCTACAGCACTAAGACAGCGTGTGAGCAAGCGGCGCTGCTACAGAGCT 1740

QY 1741 GAGGCGCCCGGACGACGACCCCTGCACAAAGACGACCCCGCTTCTGGGCTAGCATGATGGA 1800
 |||||||
 Db 1741 GAGGCGCCCGGACGACGACCCCTGCACAAAGACGACCCCGCTTCTGGGCTAGCATGATGGA 1800
 QY 1801 GGCAGCCAGCTGCCCGGATTTATGTTCTCCAGAGATCTTCATTTCCACACGCGCTC 1860
 |||||||
 Db 1801 GGCAGCCAGCTGCCCGGATTTATGTTCTCCAGAGATCTTCATTTCCACACGCGCTC 1860
 QY 1861 GGAATGCGGCTCTACGCGCATGATCTACAAAGCCCGCTTGCAGCCAGGGAAGAGA 1920
 |||||||
 Db 1861 GGAATGCGGCTCTACGCGCATGATCTACAAAGCCCGCTTGCAGCCAGGGAAGAGA 1920
 QY 1921 CCCACGCTCTCTTTTATATGAGGAGCCCGCAGAGTGCAGCTGTGATTAATCTCTTCAA 1980
 |||||||
 Db 1921 CCCACGCTCTCTTTTATATGAGGAGCCCGCAGAGTGCAGCTGTGATTAATCTCTTCAA 1980
 QY 1981 AGGCATCAAGTACTTGGCGCTCAACACACTGCGCTCCCTGGGCTAGCCCGTGTGTGAT 2040
 |||||||
 Db 1981 AGGCATCAAGTACTTGGCGCTCAACACACTGCGCTCCCTGGGCTAGCCCGTGTGTGAT 2040
 QY 2041 TGACGCGAGGGGCTCTCTGTACGACGAGGCTTCCGTTGGAAAGGGCCCTGAAAAACAAT 2100
 |||||||
 Db 2041 TGACGCGAGGGGCTCTCTGTACGACGAGGCTTCCGTTGGAAAGGGCCCTGAAAAACAAT 2100
 QY 2101 GGGCGAGGTGAGATCGAGGACGAGTGAAGGGCTGAGTTCGTTGGCGGAGATGATGG 2160
 |||||||
 Db 2101 GGGCGAGGTGAGATCGAGGACGAGTGAAGGGCTGAGTTCGTTGGCGGAGATGATGG 2160
 QY 2161 CTTTCATGACCTGAGCCGAGTGTGCTATCCATGAGCTGTCTTCTACGCGGCTTCTCTGCT 2220
 |||||||
 Db 2161 CTTTCATGACCTGAGCCGAGTGTGCTATCCATGAGCTGTCTTCTACGCGGCTTCTCTGCT 2220
 QY 2221 CATGGGCTTAATCCACAAGCCCGCAGGTTCACAGTGGCCATCGCGGCTGCCCGGTAC 2280
 |||||||
 Db 2221 CATGGGCTTAATCCACAAGCCCGCAGGTTCACAGTGGCCATCGCGGCTGCCCGGTAC 2280
 QY 2281 CGTGTGATGGCTACGACAGAGGCTACACTGAGCGCTTACATGAGAGCTCCCTGAGACAA 2340
 |||||||
 Db 2281 CGTGTGATGGCTACGACAGAGGCTACACTGAGCGCTTACATGAGAGCTCCCTGAGACAA 2340
 QY 2341 CCAGCAGCGCTATGAGCGGGCTCGTGCCCTGACAGCTGAGAGAGCTGCCCAATGAGCC 2400
 |||||||
 Db 2341 CCAGCAGCGCTATGAGCGGGCTCGTGCCCTGACAGCTGAGAGAGCTGCCCAATGAGCC 2400
 QY 2401 CAACCGCTGTATCTCTCCACAGGCTCTGAGAGAAAGCTGCACTTTTCCACACAAA 2460
 |||||||
 Db 2401 CAACCGCTGTATCTCTCCACAGGCTCTGAGAGAAAGCTGCACTTTTCCACACAAA 2460
 QY 2461 CTTCTCTGCTCTCCCACTGATCCGAGCAGGAAACTTACAGCTCCAGATCTACCCAA 2520
 |||||||
 Db 2461 CTTCTCTGCTCTCCCACTGATCCGAGCAGGAAACTTACAGCTCCAGATCTACCCAA 2520
 QY 2521 CGAGAGACACAGTATGCTGCTGCCCGAGTGGGCGAGCAGCTATGAAAGTACGTTGTGCA 2580
 |||||||
 Db 2521 CGAGAGACACAGTATGCTGCTGCCCGAGTGGGCGAGCAGCTATGAAAGTACGTTGTGCA 2580
 QY 2581 CTTTCTACAGGAATACCTGAGGCGCGGCGGATCCG 2617
 |||||||
 Db 2581 CTTTCTACAGGAATACCTGAGGCGCGGCGGATCCG 2617

RESULT 2

ABK83335 standard; cDNA; 4219 BP.

AC ABK83335;

DT 12-AUG-2002 (first entry)

DE cDNA encoding human DPRP-2 splice variant #3.

KX Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP.

KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinnesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder; gene; ss.
 OS Homo sapiens.
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001MO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Q1 S, Akinsanya KO, Riviere PJ, Junten J;
 XX
 DR WPI; 2002-44178/47.
 DR P-PSDB; AB661604.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 PS Disclosure; Page 84-85; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
 CC
 SQ Sequence 4219 BP; 908 A; 1320 C; 1190 G; 801 T; 0 other;
 Query Match 99.1%; Score 2592.4; DB 24; Length 4219;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 11 ATGGCCACACCGGACCGGACCGGACGAGCGAGCGGCGCCACAGATGACCG 70
 |||||||
 Db 367 ATGGCCACACCGGACCGGACCGGACGAGCGAGCGAGCGGCGCCACAGATGACCG 426
 QY 71 GCGCGCGCTCCAGGTGACAAAGCACTCGTGGAGCGGCTCCGAGCATATCCAGGC 130
 |||||||
 Db 427 GCGCGCGCTCCAGGTGACAAAGCACTCGTGGAGCGGCTCCGAGCATATCCAGGC 486
 QY 131 AGCCGAGATGCTGGGCTCATGTCACAAAGCGGCCCGCAGACTTCAGATTGTGACG 190
 |||||||
 Db 487 AGCCGAGATGCTGGGCTCATGTCACAAAGCGGCCCGCAGACTTCAGATTGTGACG 546
 QY 191 AAGACGATGAGTGTGGGCTCCACCGCTTACTACTGAGATGCGGAATGCCATGAGC 250
 |||||||
 Db 547 AAGACGATGAGTGTGGGCTCCACCGCTTACTACTGAGATGCGGAATGCCATGAGC 606
 QY 251 AGCCGAGAGATGCTGCTCTACTGAGATGCGGAATGCCATGAGC 310
 |||||||
 Db 607 AGCCGAGAGATGCTGCTCTACTGAGATGCGGAATGCCATGAGC 666
 QY 311 CTGCTCTGTCTGGAAGCAGATGCTGATATTTCCAGGCCACGCCACATGAGGCTC 370

|||||
Db 667 CTCTCTCTGTCGGAAGAGATGCTGGATCATTTCCAGGCCACGCCCCACCATGGGGTTC 726
OY 371 TACTCTGGAGAGAGAGCTGTAGAGGACGGAACGCCCTGGGGGCTTCGGCATACC 430
|||||
Db 727 TACTCTGGAGAGAGAGCTGTAGAGGACGGAACGCCCTGGGGGCTTCGGCATACC 786
OY 431 TCCTAGACTTCCACACGAGAGTGGCTCTCTCTCTCCAGGCCAACAGCCTCTTC 490
|||||
Db 787 TCCTAGACTTCCACACGAGAGTGGCTCTCTCTCTCCAGGCCAACAGCCTCTTC 846
OY 491 CACTGTGCGACGGCGCAAGAGCGCTTATGTGTCTCCCTATGAACCGCTGGAATC 550
|||||
Db 847 CACTGTGCGACGGCGCAAGAGCGCTTATGTGTCTCCCTATGAACCGCTGGAATC 906
OY 551 AAGACCCAGTGTCAAGGGCCCCGGATGGACCCCAAAATCTGCCCTGGCACCCCTTC 610
Db 907 AAGACCCAGTGTCAAGGGCCCCGGATGGACCCCAAAATCTGCCCTGGCACCCCTTC 966
OY 611 TTCTCTTCAATCAATAACAGCAGCTGTGGTGGCCAACTCGACAGCAGAGAGCG 670
Db 967 TTCTCTTCAATCAATAACAGCAGCTGTGGTGGCCAACTCGACAGCAGAGAGCG 1026
OY 671 CGGCTGACCTTCTGCCACCAAGGTTTATCCAAATGCTCGATGAACCCCAAGTCTGGGT 730
Db 1027 CGGCTGACCTTCTGCCACCAAGGTTTATCCAAATGCTCGATGAACCCCAAGTCTGGGT 1086
OY 731 GTGGCCACCTTGTGTCATACAGGAAGTTGACCGCTTCACTGGGTCTMGSTGTGCC 790
Db 1087 GTGGCCACCTTGTGTCATACAGGAAGTTGACCGCTTCACTGGGTCTMGSTGTGCC 1146
OY 791 ACAGCCTCTGGGAAGGTTTCAAGAGGCGCTCAAGACGCTGCGAATCTGTATGAGAAATC 850
Db 1147 ACAGCCTCTGGGAAGGTTTCAAGAGGCGCTCAAGACGCTGCGAATCTGTATGAGAAATC 1206
OY 851 GATGAGTCCGAGTGGAGGTTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGAGAGACG 910
Db 1207 GATGAGTCCGAGTGGAGGTTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGAGAGACG 1266
OY 911 GATGAGTCCGAGTACCACAGAGAGGACAGCAAGATCCAAAGATGGCTTGAATCTGGCT 970
Db 1267 GATGAGTCCGAGTACCACAGAGAGGACAGCAAGATCCAAAGATGGCTTGAATCTGGCT 1326
OY 971 GAGTCCAGACTGACAGCAGCGGCGAAGATGCTCTGACCCAGAGAAAGAGAGTGTGCAG 1030
Db 1327 GAGTCCAGACTGACAGCAGCGGCGAAGATGCTCTGACCCAGAGAAAGAGAGTGTGCAG 1386
OY 1031 CCCTTCAGCTCGCTGTTCCTCCGAGGTGGAGTACATCGCCAGGGCCGGGTGGACCCGGGAT 1090
Db 1387 CCCTTCAGCTCGCTGTTCCTCCGAGGTGGAGTACATCGCCAGGGCCGGGTGGACCCGGGAT 1446
OY 1091 GGCAGATACGCTGGGACATGTCCTGGACCGGCCCCAGAGAGTGGCTCCAGCTGTCTC 1150
Db 1447 GGCAGATACGCTGGGACATGTCCTGGACCGGCCCCAGAGAGTGGCTCCAGCTGTCTC 1506
OY 1151 CTCCTCCCGGCGCTGTTCATCCGAGCACAGAGATGAGGACGCGGCTGTGCC 1210
Db 1507 CTCCTCCCGGCGCTGTTCATCCGAGCACAGAGATGAGGACGCGGCTGTGCC 1566
OY 1211 AAGAGCTGCCCGAGAAATGTCACCGGTATGTGTACAGAGAGGTCAACCAAGTCTGG 1270
Db 1567 AAGAGCTGCCCGAGAAATGTCACCGGTATGTGTACAGAGAGGTCAACCAAGTCTGG 1626
OY 1271 ATCAATGTTCATACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGCTGTGC 1330
Db 1627 ATCAATGTTCATACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTGTGC 1686
OY 1331 TTCTCTCGGCGCAATGATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCGGTT 1390
Db 1687 TTCTCTCGGCGCAATGATGCAAGACCGGCTTCTGCCATTTGTACAAAGTCAACGCGGTT 1746
OY 1391 TTAAATCCAGAGGCTACGATTTGGAGTGGACCTTCAAGCCCGGGGAGAGATGAATTTAG 1450
|||||

Db 1747 TTAAATCCAGAGGCTACGATTTGGAGTGAACCTTCAAGCCCGGGGAGAGATTTAG 1806
OY 1451 TGCCCATTAAGGAAGAGATTTGCTGTACAGCGGTGAATGGAGGTTTGGCAGGAC 1510
Db 1807 TGCCCATTAAGGAAGAGATTTGCTGTACAGCGGTGAATGGAGGTTTGGCAGGAC 1866
OY 1511 GGCTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCCACCAAGAC 1570
Db 1867 GGCTCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGCCACCAAGAC 1926
OY 1571 ACGCGCTGAGAGACCAACCTCTACGTGTGTAGCTATGAGGCGGCGGAGATCTGACG 1630
Db 1927 ACGCGCTGAGAGACCAACCTCTACGTGTGTAGCTATGAGGCGGCGGAGATCTGACG 1986
OY 1631 CTCACACGCGCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTCGACATGTTCGTC 1690
Db 1987 CTCACACGCGCGGCTTCTCCATAGCTGTCCATGAGCCAGAACTTCGACATGTTCGTC 2046
OY 1691 AGCCACTACAGACGCTGAGACGCGCCCTGTGCTGACGCTTACAAAGTGAAGGCGGCC 1750
Db 2047 AGCCACTACAGACGCTGAGACGCGCCCTGTGCTGACGCTTACAAAGTGAAGGCGGCC 2106
OY 1751 GACGAGACCCCTGCAACAAGACGCGCCCTGTGGCTAGCATGATGAGACGACGAC 1810
Db 2107 GACGAGACCCCTGCAACAAGACGCGCCCTGTGGCTAGCATGATGAGACGACGAC 2166
OY 1811 TGCCCCCGGATTTATGTCTCTCCAGAGATCTCCATTTCCACAGCGGCTGATGTGCG 1870
Db 2167 TGCCCCCGGATTTATGTCTCTCCAGAGATCTCCATTTCCACAGCGGCTGATGTGCG 2226
OY 1871 CTCTACGGCATGTCTACAAAGCCCAAGCCCTTGCAGCCAGAGGAAGACACCCACGTC 1930
Db 2227 CTCTACGGCATGTCTACAAAGCCCAAGCCCTTGCAGCCAGAGGAAGACACCCACGTC 2286
OY 1931 CTCTTTGATATGAGAGGCCCCCGAGGTGCAAGCTGAATTAACCTTCAAAAGCATCAAG 1990
Db 2287 CTCTTTGATATGAGAGGCCCCCGAGGTGCAAGCTGAATTAACCTTCAAAAGCATCAAG 2346
OY 1991 TACTTGCGGCTCAACACACTGGCTCCCTGGGCTAGCGCTGTGTGATTTGAGGCGAG 2050
Db 2347 TACTTGCGGCTCAACACACTGGCTCCCTGGGCTAGCGCTGTGTGATTTGAGGCGAG 2406
OY 2051 GGCTCTGTGACGAGGGCTTGGGTGGAAGGGCCCTGA AAAACCAATGGGCCAGGTG 2110
Db 2407 GGCTCTGTGACGAGGGCTTGGGTGGAAGGGCCCTGA AAAACCAATGGGCCAGGTG 2466
OY 2111 GAGATGAGAGCAGGTGGAGGGGCGTGCAGTGTGGCGAGAAATGATGGCTCATCGAC 2170
Db 2467 GAGATGAGAGCAGGTGGAGGGGCGTGCAGTGTGGCGAGAAATGATGGCTCATCGAC 2526
OY 2171 CTGAGCCGAGTTGCAATCATGAGCTGTGCTTACGAGGGGCTTCTCTGCTCATGGGGCTA 2230
Db 2527 CTGAGCCGAGTTGCAATCATGAGCTGTGCTTACGAGGGGCTTCTCTGCTCATGGGGCTA 2586
OY 2231 ATCCACAAGCCCCAGGTGTTCAAGGTGCCATCGCGGGTCCACCGTCTGGATG 2290
Db 2587 ATCCACAAGCCCCAGGTGTTCAAGGTGCCATCGCGGGTCCACCGTCTGGATG 2646
OY 2291 GGCTTACGACACAGGGTACACTGAGGCTACATGAGAGTGTCCCTGAGAAACAACGACGCG 2350
Db 2647 GGCTTACGACACAGGGTACACTGAGGCTACATGAGAGTGTCCCTGAGAAACAACGACGCG 2706
OY 2351 TATGAGCGGGGTTCCGTGGCCCTGCAAGTGGAGAACTGCCCAATGAGCCCAAGCCGTTG 2410
Db 2707 TATGAGCGGGGTTCCGTGGCCCTGCAAGTGGAGAACTGCCCAATGAGCCCAAGCCGTTG 2766
OY 2411 CTATATCTCCAGGGCTTCTGAGAGAAAAGTGCACCTTTTCCACAACAACTTCTCGTC 2470
Db 2767 CTATATCTCCAGGGCTTCTGAGAGAAAAGTGCACCTTTTCCACAACAACTTCTCGTC 2826
OY 2471 TCCCACTGATCCGAGCAGAGGAACCTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2530
Db 2827 TCCCACTGATCCGAGCAGAGGAACCTTACAGCTCCAGATCTACCCCAAGAGAGACAC 2886
|||||

OY 2531 AGATTGCGTCCCGAGTGGGCGAGACATATGAATGACGTTGCTGACCTTTCTACAG 2590
 DB 2887 AGATTGCGTCCCGAGTGGGCGAGACATATGAATGACGTTGCTGACCTTTCTACAG 2946
 OY 2591 GAATACCTCTGAGC 2604
 DB 2947 GAATACCTCTGAGC 2960
 RESULT 3
 ABR83333
 ID ABR83333 standard; cDNA; 4302 BP.
 XX ABR83333;
 AC ABR83333;
 XX 12-AUG-2002 (first entry)
 DE cDNA encoding human DPRP-2 splice variant #1.
 XX
 KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KM DPP1V; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KM dyskinesia; reproductive disorder; inflammatory disorder;
 KM metabolic disorder; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 DR WPI; 2002-444178/47.
 XX
 P-Psdb; ABR61602.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 PS Disclosure; Page 78-79; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPP1V)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83332-ABR83343 encode human DPRP proteins.
 XX
 SO Sequence 4302 BP; 923 A; 1350 G; 1221 G; 808 T; 0 other:
 Query Match 99.1%; Score 2592.4; DB 24; Length 4302;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2593; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 ATGGCCACCACCGGAGCCCAAGCCGACCGAGGCGACGCGCCACAGATGACCG 70
 DB 367 ATGGCCACCACCGGAGCCCAAGCCGACCGAGGCGACGCGCCACAGATGACCG 426
 OY 71 GCCGCCCGCTTCCAGGTGCAAGAAAGCACTCGTGGAGCGGCTCCGAGCATATCAACGG 130
 DB 427 GCCGCCCGCTTCCAGGTGCAAGAAAGCACTCGTGGAGCGGCTCCGAGCATATCAACGG 486
 OY 131 AGCCGCAAGTACTCGGGGCTCATTTGTCACAAAGGCGCCGACACTTCCAGTTTGTGCAG 190
 DB 487 AGCCGCAAGTACTCGGGGCTCATTTGTCACAAAGGCGCCGACACTTCCAGTTTGTGCAG 546
 OY 191 AAGACGATGATGCTGGGCGCCGCTCCACCGGCTCTACTACTGGGAAATGCCATATGGC 250
 DB 547 AAGACGATGATGCTGGGCGCCGCTCCACCGGCTCTACTACTGGGAAATGCCATATGGC 606
 OY 251 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAGGTCGGAAGAGGCTCTG 310
 DB 607 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAGGTCGGAAGAGGCTCTG 666
 OY 311 CTGCTCTGCTCTGCGAAGCAGATGCTGATCATTTCCAGGCGCACGCGCCACATGGGCTC 370
 DB 667 CTGCTCTGCTCTGCGAAGCAGATGCTGATCATTTCCAGGCGCACGCGCCACATGGGCTC 726
 OY 371 TACTCTCGGAGGAGGAGGAGTGTGAGGAGCGGAAAGCGCTGGGGGCTTCGGCATCAC 430
 DB 727 TACTCTCGGAGGAGGAGTGTGAGGAGCGGAAAGCGCTGGGGGCTTCGGCATCAC 786
 OY 431 TCCTAGACTTCCACAGCGAGAGTGGCTTCCTCTCCAGGCGCACAGACCTCTTC 490
 DB 787 TCCTAGACTTCCACAGCGAGAGTGGCTTCCTCTCCAGGCGCACAGACCTCTTC 846
 OY 491 CACTGTGCGAGCGCGGCAAGAAAGCGCTTATGCTGTCCCTATGAACCGCTGGAATC 550
 DB 847 CACTGTGCGAGCGCGGCAAGAAAGCGCTTATGCTGTCCCTATGAACCGCTGGAATC 906
 OY 551 AAGACCAAGTGTCCAGGGCGCGGATGAGACCCCAAAATCGGCTGCGACCTGCTTC 610
 DB 907 AAGACCAAGTGTCCAGGGCGCGGATGAGACCCCAAAATCGGCTGCGACCTGCTTC 966
 OY 611 TTCTCTTCATCAATAACAGACGACCTGTGGTGGCCAAATCGAGACGCGAGGAGCGG 670
 DB 967 TTCTCTTCATCAATAACAGACGACCTGTGGTGGCCAAATCGAGACGCGAGGAGCGG 1026
 OY 671 CGGCTGACTTTCGCCACGCAAGGTTTATCCATGCTCGATGACCCCAAGTCTGGGGT 730
 DB 1027 CGGCTGACTTTCGCCACGCAAGGTTTATCCATGCTCGATGACCCCAAGTCTGGGGT 1086
 OY 731 GTGGCCACCTTGTGCATACAGGAAGTGTGACCGGCTCACTGGGTACTGGGTGCGCC 790
 DB 1087 GTGGCCACCTTGTGCATACAGGAAGTGTGACCGGCTCACTGGGTACTGGGTGCGCC 1146
 OY 791 ACAGCCTCTGGAAGGTTTCAGAGGCGCTCAAGCGGTGGAATCTGTATGAGGAATC 850
 DB 1147 ACAGCCTCTGGAAGGTTTCAGAGGCGCTCAAGCGGTGGAATCTGTATGAGGAATC 1206
 OY 851 GATGAGTCCGAGGTGAGGTTCATTCACGTCCCTCTCTCGCTAGAGAAGAGCAAGC 910
 DB 1207 GATGAGTCCGAGGTGAGGTTCATTCACGTCCCTCTCTCGCTAGAGAAGAGCAAGC 1266
 OY 911 GATCGGATGAGGTACCCGAGGACAGGACGAGCAAGATCCCAAGATGCTTGAATGCTG 970
 DB 1267 GATCGGATGAGGTACCCGAGGACAGGACGAGCAAGATCCCAAGATGCTTGAATGCTG 1326
 OY 971 GAGTTCAGAGTACAGCGCAGGCAAGATGCTTCAGCCAGAGAGAAGAGAGCTGTGTCAG 1030
 DB 1327 GAGTTCAGAGTACAGCGCAGGCAAGATGCTTCAGCCAGAGAGAAGAGAGCTGTGTCAG 1386
 OY 1031 CCCTTACGCTGCTGTTCGGAAGGTGAGTACATCGCAGGCGCGGTGACCCGGAT 1090
 DB 1387 CCCTTACGCTGCTGTTCGGAAGGTGAGTACATCGCAGGCGCGGTGACCCGGAT 1446
 OY 1091 GGCAATAGCGCTGGGCGATGTTCTGAGACGCGCCGACAGTGGCTTCAGCTGCTCCTC 1150

```

|||||
Db 1447 GCAAAATACGCTGGGGCCATGTTCTCGAGCCGCCCCAGCAGCTGGCTCCAGCTCGCTCC 1506
Oy 1151 CTCGCCCCGGGCTGTATATCCGAGACACAGAAATGAGAGACAGGGCTACCTCTGGC 1210
Db 1507 CTCGCCCCGGGCTGTATATCCGAGACACAGAAATGAGAGACAGGGCTACCTCTGGC 1566
Oy 1211 AGAGCTGCCGAGAAATGTCCAGCCGTATGTGTGTACGAGAGAGTCCACCAAGCTTGG 1270
Db 1567 AGAGCTGCCGAGAAATGTCCAGCCGTATGTGTGTACGAGAGAGTCCACCAAGCTTGG 1626
Oy 1271 ATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGACAGCTCTGC 1330
Db 1627 ATCAATGTTTCATGACATCTTCTATCCCTTCCCAATCAGAGAGAGAGACAGCTCTGC 1686
Oy 1331 TTTTCGCGGCAATGAATGCAAGACCGGCTCTGCAATTTTACAAATGCAACCGCGTT 1390
Db 1687 TTTTCGCGGCAATGAATGCAAGACCGGCTCTGCAATTTTACAAATGCAACCGCGTT 1746
Oy 1391 TTTAAATCCAGGCTACGATTGAGTGAGCCCTTCAGCCCGGGGAAGATGAATTTAAG 1450
Db 1747 TTTAAATCCAGGCTACGATTGAGTGAGCCCTTCAGCCCGGGGAAGATGAATTTAAG 1806
Oy 1451 TCGCCCATTAAGAGAGAGATTCTCTGACACAGCGTGAATGGAGGTTTGGCGAGGCAC 1510
Db 1807 TCGCCCATTAAGAGAGAGATTCTCTGACACAGCGTGAATGGAGGTTTGGCGAGGCAC 1866
Oy 1511 GGCCTCAAGATCTGGGTCATGAGAGAGACCAAGCTGTGTACTTCCAGGAGCAACAAAGAC 1570
Db 1867 GGCCTCAAGATCTGGGTCATGAGAGAGACCAAGCTGTGTACTTCCAGGAGCAACAAAGAC 1926
Oy 1571 AGCGCGCTGAGACACACCTTACGTGTGCTACGCTATGAGGCGCGCGAGATCGTACGC 1630
Db 1927 AGCGCGCTGAGACACACCTTACGTTGCTACGCTATGAGGCGCGCGAGATCGTACGC 1986
Oy 1631 CTCACACAGCGCGGCTTCTCCCATAGCTGCTCATGAGGCAACCTTGACATGTTCTGTC 1690
Db 1987 CTCACACAGCGCGGCTTCTCCCATAGCTGCTCATGAGGCAACCTTGACATGTTCTGTC 2046
Oy 1691 AGCCACTACAGAGGCTGAGAGACGCGCGCTGCGTGCAGCTGTACAAAGGTGAGCGGCCCC 1750
Db 2047 AGCCACTACAGAGGCTGAGAGACGCGCGCTGCGTGCAGCTGTACAAAGGTGAGCGGCCCC 2106
Oy 1751 GACGACGAGCCCCCTGCACAAAGACGCGCTTCTGGGCTAGCATGATGAGAGGACGACGAC 1810
Db 2107 GACGACGAGCCCCCTGCACAAAGACGCGCTTCTGGGCTAGCATGATGAGAGGACGACGAC 2166
Oy 1811 TCGCCCCGGGATTTATGTTCTCTCCAGAGATCTTCCATTTCCACACGCGCTCGATGTGCG 1870
Db 2167 TCGCCCCGGGATTTATGTTCTCTCCAGAGATCTTCCATTTCCACACGCGCTCGATGTGCG 2226
Oy 1871 CTCACGAGCATGCTACAAAGCCCCACGCGCTTGCAAGCGGAGAGAGACACCCACCGTCC 1930
Db 2227 CTCACGAGCATGCTACAAAGCCCCACGCGCTTGCAAGCGGAGAGAGACACCCACCGTCC 2286
Oy 1931 CTCCTTTGATATGAGAGCCCCAGGCTGACAGCTGTGTAATTAACCTCTTCAAAAGCATGAAG 1990
Db 2287 CTCCTTTGATATGAGAGCCCCAGGCTGACAGCTGTGTAATTAACCTCTTCAAAAGCATGAAG 2346
Oy 1991 TACTTGGCGGCTCAACACACTGGCTCTCCCTGAGGCTACGCGCTGTGTGATTCACGCGCAAG 2050
Db 2347 TACTTGGCGGCTCAACACACTGGCTCTCCCTGAGGCTACGCGCTGTGTGATTCACGCGCAAG 2406
Oy 2051 GGCCTCTGTACAGAGGCTTGGGTTGCAAGGGGCGCTGAAAAACCAAAATGGGCGCAGGTG 2110
Db 2407 GGCCTCTGTACAGAGGCTTGGGTTGCAAGGGGCGCTGAAAAACCAAAATGGGCGCAGGTG 2466
Oy 2111 GAGATCGAGAGACAGGTGAGGCGCTGCAAGTTCGTGGCCGAGAGATGAGCTTCAATCGAC 2170
Db 2467 GAGATCGAGAGACAGGTGAGGCGCTGCAAGTTCGTGGCCGAGAGATGAGCTTCAATCGAC 2526
Oy 2171 CTGAGCCGAGTTGCCATCATGGCTGTCTACGGGGGCTTCTCTCGTCAATGGGCGCTA 2230
|||||

```

```

Db 2527 CTGAGCCGAGTTGCCATCCATGCGTGTCTTACGGGGGCTTCTCTGCTCATGAGGCGCTA 2586
Oy 2231 ATCCACAAGCCCCAGGCTGTCAAGGTGGCCATGCGGGTGCCTCCGGTACCGCTGTGATG 2290
Db 2587 ATCCACAAGCCCCAGGCTGTCAAGGTGGCCATGCGGGTGCCTCCGGTACCGCTGTGATG 2646
Oy 2291 GCGTACGACAGAGGTTACACTGAGCGCTTACATGAGACGTCCCTGAGAACACACGACGCG 2350
Db 2647 GCGTACGACAGAGGTTACACTGAGCGCTTACATGAGACGTCCCTGAGAACACACGACGCG 2706
Oy 2351 TATGAGGGGGGTTCCGTGCGCCCTGACGTGAGAGCTGCCCCAATGAGCCCAACCGCTTG 2410
Db 2707 TATGAGGGGGGTTCCGTGCGCCCTGACGTGAGAGCTGCCCCAATGAGCCCAACCGCTTG 2766
Oy 2411 CTTATCTCCACAGGCTTCCCTGAGCAAAAACGCACTTTTTCACACAACCTTCTCTGTC 2470
Db 2767 CTTATCTCCACAGGCTTCCCTGAGCAAAAACGCACTTTTTCACACAACCTTCTCTGTC 2826
Oy 2471 TCCCAACTGATCCGAGACAGAGGAAACCTTACACAGCTCCAGATCTACCCCAAGAGACAC 2530
Db 2827 TCCCAACTGATCCGAGACAGAGGAAACCTTACACAGCTCCAGATCTACCCCAAGAGACAC 2886
Oy 2531 AGTATTCGCTGCGCGAGTGGGGGAGACACTATGAAGTCAGTGTGCTGACTTCTTACAG 2590
Db 2887 AGTATTCGCTGCGCGAGTGGGGGAGACACTATGAAGTCAGTGTGCTGACTTCTTACAG 2946
Oy 2591 GAATACCTCTGAGC 2604
Db 2947 GAATACCTCTGAGC 2960

```

```

RESULT 4
AAD38954
ID AAD38954 standard: cDNA: 3024 BP.
XX
AC AAD38954;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human dipeptidyl peptidase 9 (DPP9) cDNA.
XX
KW Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis;
KW autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;
KW grief rejection; antidiabetic; antinflammatory; immunosuppressive;
KW antiviral; enzyme; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2910
FT /tag= a
FT /product= "Human DPP9 protein"
FT /transl_except= (pos: 1120..1122, aa:Gln)
FT /note= "CDS does not include start codon"
FT /partial
XX
MO200234900-A1.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-A001388.
XX
PR 27-OCT-2000; 2000AU-0001078.
XX
PA (UNSY ) UNIV SYDNEY.
XX
PI Abbott CA, Gorrell MD;
XX
PI WPI: 2002-454646/48.
XX
DR P-PSDB: AAE24168.
XX
XX New dipeptidyl peptidase (DPP) peptides, useful for screening
PT inhibitors of DPP catalytic activity, which may be employed to treat

```

PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
rejection and HIV infection -
XX
PS Example; Fig 4; 91pp; English.

CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
CC polynucleotides encoding such proteins. The DPP peptides are useful for
CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
CC rejection and HIV (human immuno deficiency virus) infection. The present
CC sequence is human DPP9 cDNA.

XX Sequence 3024 BP; 624 A; 973 C; 875 G; 552 T; 0 other;

Query Match 98.9%; Score 2587.6; DB 24; Length 3024;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2550; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 ATGGCCACACCGGAGCCCGACGCGGACGAGCGACGACGCGCGGACGATGACCG 70
DB 319 ATGGCCACACCGGAGCCCGACGCGGACGAGCGGACGAGCGCGGACGATGACCG 378
QY 71 GCGCCCGCTTCCAGGTGCGAAGCACTGCTGGAGCGGCTCCGGAGCATTCACGCG 130
DB 379 GCGCCCGCTTCCAGGTGCGAAGCACTGCTGGAGCGGCTCCGGAGCATTCACGCG 438
QY 131 AGCGGCAAGTCTGGGCGCTCATTTGCAACAAGCGCGCGCGCGGACGATTCAG 190
DB 439 AGCGGCAAGTCTGGGCGCTCATTTGCAACAAGCGCGCGCGCGGACGATTCAG 498
QY 191 AAGAGGATGATCTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 250
DB 499 AAGAGGATGATCTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 558
QY 251 AGCGGAGAGATCT 310
DB 559 AGCGGAGAGATCT 618
QY 311 CTGCT 370
DB 619 CTGCT 678
QY 371 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAAGCGCTGGGGGCTTTCGGCATACC 430
DB 679 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAAAGCGCTGGGGGCTTTCGGCATACC 738
QY 431 TCTTAGACTTCCACAGCGAGAGTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 490
DB 739 TCTTAGACTTCCACAGCGAGAGTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 798
QY 491 CACTGTGCGAGCGGCGGCAAGAGGCTTATGATGTGCTCCCTATGAAAGCGTGGAAATC 550
DB 799 CACTGTGCGAGCGGCGGCAAGAGGCTTATGATGTGCTCCCTATGAAAGCGTGGAAATC 858
QY 551 AAGACCAAGTCTCAGGCGCGCGGATGAGACCCCAAAATCTGCCTGCGACGCTGCTTC 610
DB 859 AAGACCAAGTCTCAGGCGCGCGGATGAGACCCCAAAATCTGCCTGCGACGCTGCTTC 918
QY 611 TTCTCTCTATCAATAACAGCGACCTGTGGTGCCCAATCGAGACAGCGAGAGCGG 670
DB 919 TTCTCTCTATCAATAACAGCGACCTGTGGTGCCCAATCGAGACAGCGAGAGCGG 978
QY 671 CGGCTGACCTTGTGCAACCAAGGTTTATCCAAATGCTCTGATGACCCCAAGTCTGGCGGT 730
DB 979 CGGCTGACCTTGTGCAACCAAGGTTTATCCAAATGCTCTGATGACCCCAAGTCTGGCGGT 1038
QY 731 GTGGCCACCTTCTCATACAGAGAGTTCGACGCTTCACTGGTACTGTGGTGCGCC 790
DB 1039 GTGGCCACCTTCTCATACAGAGAGTTCGACGCTTCACTGGTACTGTGGTGCGCC 1098
QY 791 ACAGCTCTCTGGAAGGTTAGAGGCGCTTAAGAAGCTGGGAATCTGTATGAGGAATC 850
DB 1099 ACAGCTCTCTGGAAGGTTAGAGGCGCTTAAGAAGCTGGGAATCTGTATGAGGAATC 1158

QY 851 GATGAGTCCGAGGAGGAGGATTCATTCAGTCCCTCTCTGCGCTAGAGAAGAGACAGC 910
DB 1159 GATGAGTCCGAGGAGGAGGATTCATTCAGTCCCTCTCTGCGCTAGAGAAGAGACAGC 1218
QY 911 GACTCGTATGATGATCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 970
DB 1219 GACTCGTATGATGATCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1278
QY 971 GAGTTCAGACTGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1030
DB 1279 GAGTTCAGACTGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1338
QY 1031 CCGTTCAGCTGCGGTTCGGAAGGTGAGTACATCCGAGGCGCGGCGGCGGCGGCGGAT 1090
DB 1339 CCGTTCAGCTGCGGTTCGGAAGGTGAGTACATCCGAGGCGCGGCGGCGGCGGCGGAT 1398
QY 1091 GGCMAATAGCGCTGGGCGATGTTCTTGAGCGGCGCGGCGGAGAGTGCTCCAGCTCTCTC 1150
DB 1399 GGCMAATAGCGCTGGGCGATGTTCTTGAGCGGCGCGGCGGAGAGTGCTCCAGCTCTCTC 1458
QY 1151 CTCCCGCGGCGCTGTTTCAATCCGAGCAGAGAGATGAGAGGAGCGGCTAGCCTTGCC 1210
DB 1459 CTCCCGCGGCGCTGTTTCAATCCGAGCAGAGAGATGAGAGGAGCGGCTAGCCTTGCC 1518
QY 1211 AGAGCTGTCCCGAGGATGTCCAGCCGTATGTGTGATGAGAGGAGGAGGAGGAGGAG 1270
DB 1519 AGAGCTGTCCCGAGGATGTCCAGCCGTATGTGTGATGAGAGGAGGAGGAGGAGGAG 1578
QY 1271 ATCAATGTTATGATCATCTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1330
DB 1579 ATCAATGTTATGATCATCTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1638
QY 1331 TTTCTCCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390
DB 1639 TTTCTCCGCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1698
QY 1391 TTTAAATCCGAGGCGTACGANTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1450
DB 1699 TTTAAATCCGAGGCGTACGANTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1758
QY 1451 TGCCCGCATTAAGAGAGATGCTGTGACAGGCGGTAATGAGGAGGAGGAGGAGGAGG 1510
DB 1759 TGCCCGCATTAAGAGAGATGCTGTGACAGGCGGTAATGAGGAGGAGGAGGAGGAGG 1818
QY 1511 GGCTCCAGATCTGGGTCATGAGAGAGCAAGCTGCTTCTTCCAGGCGACCAAGGAGC 1570
DB 1819 GGCTCCAGATCTGGGTCATGAGAGAGCAAGCTGCTTCTTCCAGGCGACCAAGGAGC 1878
QY 1571 AGCGCGCTGAGAGCAACGCTCTACGTGTGATGAGTATGAGGCGGCGGCGGAGTCTACGC 1630
DB 1879 AGCGCGCTGAGAGCAACGCTCTACGTGTGATGAGTATGAGGCGGCGGCGGAGTCTACGC 1938
QY 1631 CTCACACAGCGCGGCTTCTCCATAGCTGTGCTCATGAGCGAGCAACTCTGCATGTTGCTC 1690
DB 1939 CTCACACAGCGCGGCTTCTCCATAGCTGTGCTCATGAGCGAGCAACTCTGCATGTTGCTC 1998
QY 1691 AGCCACTACAGAGCGTGAACAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1750
DB 1999 AGCCACTACAGAGCGTGAACAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2058
QY 1751 GAGGAGAGCGCGCTGCAAGAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1810
DB 2059 GAGGAGAGCGCGCTGCAAGAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2118
QY 1811 TGCCCGCGGATTAATGTTCTCCAGAGATCTTCCATTTCCACAGCGCGCTGAGATGTCGG 1870
DB 2119 TGCCCGCGGATTAATGTTCTCCAGAGATCTTCCATTTCCACAGCGCGCTGAGATGTCGG 2178
QY 1871 CTCTAGGCGATGATCTTCAAGAGCGCGCTTGCAGCGGAGGAGGAGGAGGAGGAGGAGG 1930
DB 2179 CTCTAGGCGATGATCTTCAAGAGCGCGCTTGCAGCGGAGGAGGAGGAGGAGGAGGAGG 2238

QY 1931 CTCCTTTGATATGAGAGCCCCAGGTGCAGCTGTGTAATACTCTTCAAAAGCATCAAG 1990
 DB 2239 CTCCTTTGATATGAGAGCCCCAGGTGCAGCTGTGTAATACTCTTCAAAAGCATCAAG 2298
 QY 1991 TACTTCGGCTCAACACACTGGGCTCCCTGAGGTACGCCCTGGTTGTGATGACGGCAGG 2050
 DB 2299 TACTTCGGCTCAACACACTGGGCTCCCTGAGGTACGCCCTGGTTGTGATGACGGCAGG 2358
 QY 2051 GCGTCTGTACAGAGAGGCTTCGTTTGAAGGGGCTGTGAAAACCAATAGGCGCAGGTG 2110
 DB 2359 GCGTCTGTACAGAGAGGCTTCGTTTGAAGGGGCTGTGAAAACCAATAGGCGCAGGTG 2418
 QY 2111 GAGATCGAGAGCAGGTGAGAGGCTGTGACATTCGTGTGGCCGAGAGATGATGCTTCATCGAC 2170
 DB 2419 GAGATCGAGAGCAGGTGAGAGGCTGTGACATTCGTGTGGCCGAGAGATGATGCTTCATCGAC 2478
 QY 2171 CTGAGCCGAGTTGCCATTCATGGCTGTGCTTCTGCTGCTCATGAGGCGCTA 2230
 DB 2479 CTGAGCCGAGTTGCCATTCATGGCTGTGCTTCTGCTGCTCATGAGGCGCTA 2538
 QY 2231 ATCCACAAGCCCCAGGTGTTCAGAGTGGCCATGCGGGGTGCCCGGTCAACCGTCTGGATG 2290
 DB 2539 ATCCACAAGCCCCAGGTGTTCAGAGTGGCCATGCGGGGTGCCCGGTCAACCGTCTGGATG 2598
 QY 2291 GCGTACGACACAGAGGTACACTGAGCCCTACATGAGCGTCCCTGAGAACCAACACAGCAGGC 2350
 DB 2599 GCGTACGACACAGAGGTACACTGAGCCCTACATGAGCGTCCCTGAGAACCAACACAGCAGGC 2658
 QY 2351 TATGAGGCGGGTTCGTTGCGGCTGCGACGTGGAGAGCTGCGCAATGAGCCCAACCGCGTTG 2410
 DB 2659 TATGAGGCGGGTTCGTTGCGGCTGCGACGTGGAGAGCTGCGCAATGAGCCCAACCGCGTTG 2718
 QY 2411 CTTATCTCTCCACGGCTTCTCTGAGAGAAAGCTGCTATTTTCCACACAACCTTCTCTGTC 2470
 DB 2719 CTTATCTCTCCACGGCTTCTCTGAGAGAAAGCTGCTATTTTCCACACAACCTTCTCTGTC 2778
 QY 2471 TCCCAACTGATCCGAGAGAGGAAACCTTACCAAGCTCCAGATCTTACCCCAACGAGAGAC 2530
 DB 2779 TCCCAACTGATCCGAGAGAGGAAACCTTACCAAGCTCCAGATCTTACCCCAACGAGAGAC 2838
 QY 2531 AGTATTCGCTGGCCGAGTGGGGGCGAGCACTATGAAAGTCACTGTGCTGCACTTCTACAG 2590
 DB 2839 AGTATTCGCTGGCCGAGTGGGGGCGAGCACTATGAAAGTCACTGTGCTGCACTTCTACAG 2898
 QY 2591 GAATACCTCTGAGC 2604
 DB 2899 GAATACCTCTGAGC 2912
 RESULT 5
 ABR83339
 ID ABR83339 standard; cDNA; 4180 BP.
 AC ABR83339;
 DE 12-AUG-2002 (first entry)
 XX cDNA encoding human DPRP-2 splice variant #7.
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyslexia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 OS Homo sapiens.
 PN MO200231134-A2.
 PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US31874.
 XX 12-OCT-2000; 2000US-240117P.
 XX (FERR) FERRING BV.
 PA Q1 S, Akinsanya KO, Riviere PJ, Junien J;
 PI MPI: 2002-444178/47.
 DR P-PSDB: A861608.
 XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 PS Disclosure; Page 97-98; 113pp; English.
 XX The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyslexias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABR83322-ABR83343 encode human DPRP proteins.
 XX Sequence 4180 BP; 898 A; 1312 C; 1178 G; 792 T; 0 other:

Query Match 95.7%; Score 2504.4; DB 24; Length 4180;
 Best local Similarity 98.5%; Pred. No. 0;
 Matches 2554; Conservative 0; Mismatches 1; Indels 39; Gaps 1;

QY 11 ATGGCCACACCGGGAGCCCAACGCGGACCGAGGAGCGACGCCGACAGATGACCCG 70
 DB 367 ATGGCCACACCGGGAGCCCAACGCGGACCGAGGAGCGACGCCGACAGATGACCCG 426
 QY 71 GCCGCCGCTTCAGGTGAGAACCACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 130
 DB 427 GCCGCCGCTTCAGGTGAGAACCACTGCTGGAGCGGCTCCGAGCATCATCCAGCGC 486
 QY 131 AGCCGCAAGTACTCGGGCTCTCATTTCAACAGAGCGCCGACAGACTTCCAGTTTGCGAG 190
 DB 487 AGCCGCAAGTACTCGGGCTCTCATTTCAACAGAGCGCCGACAGACTTCCAGTTTGCGAG 546
 QY 191 AAGACGATGAGTCTGGGCGCCCACTCCACGCGCTTACTACTCTGGGAATGCAATGAGC 250
 DB 547 AAGACGATGAGTCTGGGCGCCCACTCCACGCGCTTACTACTCTGGGAATGCAATGAGC 606
 QY 251 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAGGTCCGGAAGAGGCTG 310
 DB 607 AGCCGAGAGAACTCCCTCTCTCTACTCTGAGATTCCCAAGAGGTCCGGAAGAGGCTG 666
 QY 311 CTGCTCTCTGCTCGGAAGCAGATGCTGATCATTTCCAGGCCACGCCCCACCATGGGGTG 370
 DB 667 CTGCTCTCTGCTCGGAAGCAGATGCTGATCATTTCCAGGCCACGCCCCACCATGGGGTG 726
 QY 371 TACTTCGGAGAGAGAGTGTGCTGAGGAGCGGAAAGCGCTGGGGTCTTGGCATCAC 430
 DB 727 TACTTCGGAGAGAGAGTGTGCTGAGGAGCGGAAAGCGCTGGGGTCTTGGCATCAC 786
 QY 431 TCTTAGACTTCACAGCAGAGAGTGGCTCTTCTCTTCCAGGCGAGCAACAGCCTCTTC 490
 DB 787 TCTTAGACTTCACAGCAGAGAGTGGCTCTTCTCTTCCAGGCGAGCAACAGCCTCTTC 846
 QY 491 CACTGTGCGAGCGCGGCAAGAGCGCTTCTGAGTGTCCCTATGAAACCGCTGGAATC 550

Db 847 CACTGCCGAGAGCGCGCAAGAACGGCTTCATGTTGTCCTTATGAAACCGCTGAAATC 906
 QY 551 AAGACCCAGTGCATAGAGGCGCCGGATGAGACCCAAATGTCGCCCTGCGACCTGCTTC 610
 Db 907 AAGACCCAGTGCATAGAGGCGCCGGATGAGACCCAAATGTCGCCCTGCGACCTGCTTC 966
 QY 611 TTCTCTTCATCAATTAACAGCAGACCTGTGGTGGCCAAATGAGAGAGAGAGAGAGAG 670
 Db 967 TTCTCTTCATCAATTAACAGCAGACCTGTGGTGGCCAAATGAGAGAGAGAGAGAGAG 1026
 QY 671 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGTCCTGATGAGACCCCAAGTCTCGGGT 730
 Db 1027 CGGCTGACCTTCTGCCCAAGGTTTATCCAAATGTCCTGATGAGACCCCAAGTCTCGGGT 1086
 QY 731 GTGGCCACTTGTGCTATTAAGAGAAAGTTCGACCGCTTCACTGGGTACTGGTGGTCCCC 790
 Db 1087 GTGGCCACTTGTGCTATTAAGAGAAAGTTCGACCGCTTCACTGGGTACTGGTGGTCCCC 1146
 QY 791 ACAGCCCTCTGGGAAGTTTCAGAGGCTCAAGAGCGCTGGAATCCTGTATGAGAAAGTTC 850
 Db 1147 ACAGCCCTCTGGGAAGTTTCAGAGGCTCAAGAGCGCTGGAATCCTGTATGAGAAAGTTC 1206
 QY 851 GATGATCCGAGGTGAGGTCAATTCACGTCCCTCTCTGCGCTAGAGAAAGAGAGAGAG 910
 Db 1207 GATGATCCGAGGTGAGGTCAATTCACGTCCCTCTCTGCGCTAGAGAAAGAGAGAGAG 1266
 QY 911 GACTGCTATCGGTATACCCCAAGAGAGAGAGAGAGAGATCCCAAGATTCCTTGAATCTGGCT 970
 Db 1267 GACTGCTATCGGTATACCCCAAGAGAGAGAGAGAGAGATCCCAAGATTCCTTGAATCTGGCT 1326
 QY 971 GAGTTCCAGACTGACAGAGAGAGAGAGATGCTGACAGCAGAGAGAGAGAGAGAGAGAG 1030
 Db 1327 GAGTTCCAGACTGACAGAGAGAGAGAGAGATGCTGACAGCAGAGAGAGAGAGAGAGAGAG 1386
 QY 1031 CCCCTTACGTCGCTGTTCCCGAAAGTGAAGTACATCCAGAGGCGCGGTGAGACCCGGAGT 1090
 Db 1387 CCCCTTACGTCGCTGTTCCCGAAAGTGAAGTACATCCAGAGGCGCGGTGAGACCCGGAGT 1446
 QY 1091 GCGAAATAGCGCTGGGCGCAATGTTCTGAGACCGGCGCCAGAGAGAGAGAGAGAGAGAG 1150
 Db 1447 GCGAAATAGCGCTGGGCGCAATGTTCTGAGACCGGCGCCAGAGAGAGAGAGAGAGAGAG 1506
 QY 1151 CTCGCCCGGCGCTGTTCAATCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
 Db 1507 CTCGCCCGGCGCTGTTCAATCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
 QY 1211 AGAGCTGTCCCGAGAGATGTCAGCCGATGTTGTTGTTAGAGAGAGAGAGAGAGAGAGAG 1270
 Db 1567 AGAGCTGTCCCGAGAGATGTCAGCCGATGTTGTTGTTAGAGAGAGAGAGAGAGAGAGAG 1626
 QY 1271 ATCAATGTTCAATGACATCTTTCTATCCCTTCCCCCAATCAGAGAGAGAGAGAGAGAGAG 1330
 Db 1627 ATCAATGTTCAATGACATCTTTCTATCCCTTCCCCCAATCAGAGAGAGAGAGAGAGAGAG 1686
 QY 1331 TTCTCTCGGCGCAATGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
 Db 1687 TTCTCTCGGCGCAATGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
 QY 1391 TTAATAATCCAGAGGCTACGATTTGAGTGAAGCCCTTACAGCCCGGAGAGAGAGATTAAG 1450
 Db 1747 TTAATAATCCAGAGGCTACGATTTGAGTGAAGCCCTTACAGCCCGGAGAGAGATTAAG 1806
 QY 1451 TGCCCATTTAAGAAAGATTTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1510
 Db 1807 TGCCCATTTAAGAAAGATTTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
 QY 1511 GGCTCCAAGATCTGGGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1570
 Db 1867 GGCTCCA-----AGGCGACCAAGAGAG 1887
 QY 1571 ACGCCGCTGAGAGACCACTTAAGTGTGACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1630
 Db 1888 ACGCCGCTGAGAGACCACTTAAGTGTGACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1947

QY 1631 CTCACAGCGCGCGGCTTCCCATAGCTGCTCCATGAGGACAGAACTTGACATGTTGCTC 1690
 Db 1948 CTCACAGCGCGGCTTCCCATAGCTGCTCCATGAGGACAGAACTTGACATGTTGCTC 2007
 QY 1691 AGCAGCTACAGAGAGCTGAGCAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1750
 Db 2008 AGCAGCTACAGAGAGCTGAGCAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2067
 QY 1751 GAGCAGACCCCTGCAAG 1810
 Db 2068 GAGCAGACCCCTGCAAG 2127
 QY 1811 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTCGGATGTCGG 1870
 Db 2128 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGGCTCGGATGTCGG 2187
 QY 1871 CTCACGAGATGATTAACAAGCCCGCAGGCTTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAG 1930
 Db 2188 CTCACGAGATGATTAACAAGCCCGCAGGCTTGCAGCCAGAGAGAGAGAGAGAGAGAGAGAG 2247
 QY 1931 CTCCTTGTATGAGAGGCGCCCGCAGGTGACGCTGAGTAATTAATCTCTTCAAGGATCAAG 1990
 Db 2248 CTCCTTGTATGAGAGGCGCCCGCAGGCTGAGTAATTAATCTCTTCAAGGATCAAG 2307
 QY 1991 TACTTGGGCTCAACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2050
 Db 2308 TACTTGGGCTCAACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2367
 QY 2051 GGCTCTGTACAGAGAGGCTTGGTTCTGAAAGAGGAGCCCTGAAACCAATAGGAGAGAGT 2110
 Db 2368 GGCTCTGTACAGAGAGGCTTGGTTCTGAAAGAGGAGCCCTGAAACCAATAGGAGAGAGT 2427
 QY 2111 GAGATCAG 2170
 Db 2428 GAGATCAG 2487
 QY 2171 CTGAGCGAGAGTCCCATTCATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2230
 Db 2488 CTGAGCGAGAGTCCCATTCATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2547
 QY 2231 ATCCACAGCGCGGCTGTTCAAGAGTGGCCATGCGGAGTGCCTCGGCTCACCGTCTGGAGT 2290
 Db 2548 ATCCACAGCGCGGCTGTTCAAGAGTGGCCATGCGGAGTGCCTCGGCTCACCGTCTGGAGT 2607
 QY 2291 GCTTACGACAGAGAGTACACTGAGCGCTACATGAGAGTCCCTGAGAAACACAGCAGCGC 2350
 Db 2608 GCTTACGACAGAGAGTACACTGAGCGCTACATGAGAGTCCCTGAGAAACACAGCAGCGC 2667
 QY 2351 TATGAGGCGGAGTCCGTTGGGCGCTGACAGTGGAGAAAGCTGCCAATGAGGCCAAGCGCTTG 2410
 Db 2668 TATGAGGCGGAGTCCGTTGGGCGCTGACAGTGGAGAAAGCTGCCAATGAGGCCAAGCGCTTG 2727
 QY 2411 CTATCTCTCCAGGCTTCTGAG 2470
 Db 2728 CTATCTCTCCAGGCTTCTGAG 2787
 QY 2471 TCCCAACTGATCCGAGCAG 2530
 Db 2788 TCCCAACTGATCCGAGCAG 2847
 QY 2531 AGTATTCGCTGCGCGAGTCCGAG 2590
 Db 2848 AGTATTCGCTGCGCGAGTCCGAG 2907
 QY 2591 GAATACCTCTGAGC 2604
 Db 2908 GAATACCTCTGAGC 2921

RESULT 6
 ABR83338
 ID ABR83338 standard; cDNA; 4263 BP.

D	b	1627	ATCAATGTTCAATGACATCTTTATTCCTTCCTCCCAATACAGAGGAGAGACAGACTCTGC	1686
Q	y	1331	TTTTCTCGCGCCATGATGACAAGACCGGCTTCTGCATTTGTACAAGTCACCGCGTT	1390
D	b	1687	TTTTCTCGCGCCATGATGACAAGACCGGCTTCTGCATTTGTACAAGTCACCGCGTT	1746
Q	y	1391	TTAAATATCCAGGGCTACAGATTGAGTAGGACCTTCACGCCCGGGGAAGATGATTTAA	1450
D	b	1747	TTAAATATCCAGGGCTACAGATTGAGTAGGACCTTCACGCCCGGGGAAGATGATTTAA	1806
Q	y	1451	TGCCCCATTAAAGAAAGATGTGCTGCACACGGGTGAATGGGAGGTTTGGCAGGGAC	1510
D	b	1807	TGCCCCATTAAAGAAAGATGTGCTGCACACGGGTGAATGGGAGGTTTGGCAGGGAC	1866
Q	y	1511	GGCTCCAAAGATCTGGGTTCAATGAGAGACCAAGCTGATGTAATTCACAGGGACCAAGAC	1570
D	b	1867	GGCTCCAAAGATCTGGGTTCAATGAGAGACCAAGCTGATGTAATTCACAGGGACCAAGAC	1887
Q	y	1571	ACGCGCGTGGAGACACACCTCTACGTGTACGCTATGAGGGCGCGCGAGATGTTACG	1630
D	b	1888	ACGCGCGTGGAGACACACCTCTACGTGTACGCTATGAGGGCGCGCGAGATGTTACG	1947
Q	y	1631	CTCACACAGCGCGGGCTTCTCCCATPACTCTCTCCATGAGCAGAACTTGTGACATGTCGTC	1690
D	b	1948	CTCACACAGCGCGGGCTTCTCCCATPACTCTCTCCATGAGCAGAACTTGTGACATGTCGTC	2007
Q	y	1691	AGGCATCTACAGACAGCGTAGACAGCGCGCTCGTGCAGCTTACAACTGACAGCGGCC	1750
D	b	2008	AGGCATCTACAGACAGCGTAGACAGCGCGCTCGTGCAGCTTACAACTGACAGCGGCC	2067
Q	y	1751	GAGCAGACCCCTCGACACAGACGCCCCGCTTCTGGGCTAGCATGATGAGGACGACG	1810
D	b	2068	GAGCAGACCCCTCGACACAGACGCCCCGCTTCTGGGCTAGCATGATGAGGACGACG	2127
Q	y	1811	TGCCCCCGGATATGTTCCTCCAGAGATCTTCACATTCACACGCGCTCGATGTGGG	1870
D	b	2128	TGCCCCCGGATATGTTCCTCCAGAGATCTTCACATTCACACGCGCTCGATGTGGG	2187
Q	y	1871	CTCTACGCGCATGATCTCAAGAGCCCAAGCGCTTGCAGCCAGGAAAGACACCCACGCTC	1930
D	b	2188	CTCTACGCGCATGATCTCAAGAGCCCAAGCGCTTGCAGCCAGGAAAGACACCCACGCTC	2247
Q	y	1931	CTCTTTGTATATGAGAGCCCAAGGTGCAGCTGTGTGAATACCTCTCAAAAGCATCAAG	1990
D	b	2248	CTCTTTGTATATGAGAGCCCAAGGTGCAGCTGTGTGAATACCTCTCAAAAGCATCAAG	2307
Q	y	1991	TACTTGGGGCTCAACACACACTGCGCTCCCTGGGCTACGCGCGTGTGTATTTAGCGGACG	2050
D	b	2308	TACTTGGGGCTCAACACACACTGCGCTCCCTGGGCTACGCGCGTGTGTATTTAGCGGACG	2367
Q	y	2051	GGCTCTGTGACGAGAGGGCTTGGTTTCAAGAGGGCCCTCAAAAACCAATGGGCGCAAGTG	2110
D	b	2368	GGCTCTGTGACGAGAGGGCTTGGTTTCAAGAGGGCCCTCAAAAACCAATGGGCGCAAGTG	2427
Q	y	2111	GAGATCGAGACCAAGGTGAGAGGGCTTGCAGATTGCTGGCCAGAGATATGCTTCATCGAC	2170
D	b	2428	GAGATCGAGACCAAGGTGAGAGGGCTTGCAGATTGCTGGCCAGAGATATGCTTCATCGAC	2487
Q	y	2171	CTGACGCGATGTGCTATCCTATGCTGTCTACGGGGCTTCTCTCGCTATGAGGGCTA	2230
D	b	2488	CTGACGCGATGTGCTATCCTATGCTGTCTACGGGGCTTCTCTCGCTATGAGGGCTA	2547
Q	y	2231	ATTCACAAGCCCCAGGTGTCAAGGTGGCCATGGGGGTGCCCCGGTCAACGCTTGAGTG	2290
D	b	2548	ATTCACAAGCCCCAGGTGTCAAGGTGGCCATGGGGGTGCCCCGGTCAACGCTTGAGTG	2607
Q	y	2291	GCCTACGACACAGGGTCACTGAGCGCTACATGAGAGCTCCTGAGAACAAACAGCAGCGC	2350
D	b	2608	GCCTACGACACAGGGTCACTGAGCGCTACATGAGAGCTCCTGAGAACAAACAGCAGCGC	2667
Q	y	2351	TATGAGGCGGTTCTCGTGCGCTCTGCAGCTGTGAGAGAGCTGCCAATGAGCCCAACGCTTG	2410
D	b	2668	TATGAGGCGGTTCTCGTGCGCTCTGCAGCTGTGAGAGAGCTGCCAATGAGCCCAACGCTTG	2727

Oy	2411	CTTATCTCCACGSGCTTCCTGGAGCAGAAACGTCACTTTTTCACACAACACTTCTCGTC	2470
Db	2728	CTTATCTCCACGSGCTTCCTGGAGCAGAAACGTCACTTTTTCACACAACACTTCTCGTC	2787
Oy	2471	TGCCAACGATGCCAGGAGGAACACCTTACCAGCTCCAGATCCACCCCAGAAGAGAAC	2530
Db	2788	TGCCAACGATGCCAGGAGGGAACCTTACCAGCTCCAGATCCACCCCAGAAGAGAAC	2847
Oy	2531	AGTATCGCTGCCCCCGAGTGCGGCGAGCACTATGAAGTCACGTTGCTGCACCTTCTACAG	2590
Db	2848	AGTATCGCTGCCCCCGAGTGCGGCGAGCACTATGAAGTCACGTTGCTGCACCTTCTACAG	2907
Oy	2591	GAAATACCTCTGAGC	2604
Db	2908	GAAATACCTCTGAGC	2921
<hr/>			
RESULT 7			
AAD38957	ID	AAD38957 standard; DNA:	2495 BP.
XX	AC	AAD38957:	
XX	DF	23-SEP-2002	(first entry)
XX	DE	Human dipeptidyl peptidase 4 (DPP4)-like 2 DNA.	
XX	KM	Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis	
XX	KW	autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;	
KW	KT	graft rejection; antidiabetic; antiinflammatory; immunosuppressive;	
XX	OS	antiviral; enzyme; gene; ds.	
XX	OS	Homo sapiens.	
FH	Key	Location/Qualifiers	
FT	CDS	1..2493	
FT		/frag= a	
FT		/product= "Human DPP4-like 2 protein"	
FT		/transl_except= (pos: 703..705, aa:Gln)	
FT		/note= "CDS does not include start codon"	
XX	PN	WO200234900-A1.	
XX	PD	02-MAY-2002.	
XX	PF	29-OCT-2001; 2001WO-AU01388.	
XX	PR	27-OCT-2000; 2000AU-0001078.	
XX	PA	(UNSY) UNIV SYDNEY.	
XX	PI	Abdolt CA, Gorrell MD:	
DR		WPI: 2002-454646/48.	
XX		P-PsDB: AAE24171.	
XX	PT	New dipeptidyl peptidase (DPP) peptides, useful for screening	
PT		inhibitors of DPP catalytic activity, which may be employed to treat	
PT		e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft	
XX		rejection and HIV Infection -	
PS		Disclosure: Page 86-88; 91pp; English.	
XX		The present invention relates to dipeptidyl peptidase (DPP) proteins and	
CC		polynucleotides encoding such proteins. The DPP peptides are useful for	
CC		screening inhibitors of DPP catalytic activity. The inhibitors are usefu	
CC		for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft	
CC		rejection and HIV (human immuno deficiency virus) infection. The present	
CC		sequence is human DPP4-like 2 DNA.	
XX		Sequence 2495 BP; 535 A; 783 C; 696 G; 481 T; 0 other;	

Query Match 95.1%; Score 2488.6; DB 24; Length 2495;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2491; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 110 CTCGGAGCATCCAGGAGCGGCAATGATCTCGGGCCCTCATGTCACAAAGGCGCC 169
Db 1 CTCGGAGCATCCAGGAGCGGCAATGATCTCGGGCCCTCATGTCACAAAGGCGCC 60

QY 170 CAGACTTCCAGTTTGCAGAAAGAGATGATCTGGGCCCACTCCACCGCTCTAC 229
Db 61 CAGACTTCCAGTTTGCAGAAAGAGATGATCTGGGCCCACTCCACCGCTCTAC 120

QY 230 TACCTGGGAATGCCATATGSCAGCCGAGAACTCCCTCTCTACTCTGAGATTTCCAA 289
Db 121 TACCTGGGAATGCCATATGSCAGCCGAGAACTCCCTCTCTACTCTGAGATTTCCAA 180

QY 290 AAGTCCGGAAGAGAGTCTGCTCTCTCTCTCTCTGGAAGAGATGCTGATTTCCAG 349
Db 181 AAGTCCGGAAGAGAGTCTGCTCTCTCTCTCTCTGGAAGAGATGCTGATTTCCAG 240

QY 350 GCCACGCCCCACCATGAGGAGTCTACTCTCGGAGAGAGAGTCTGAGGAGCGGAACGC 409
Db 241 GCCACGCCCCACCATGAGGAGTCTACTCTCGGAGAGAGAGTCTGAGGAGCGGAACGC 300

QY 410 CTGGGGCTTTCGGCATCACCTCTACGACTTCCACAGCAGAGATGAGCTTCTCTTC 469
Db 301 CTGGGGCTTTCGGCATCACCTCTACGACTTCCACAGCAGAGATGAGCTTCTCTTC 360

QY 470 CAGGCCAGCAAGCCCTTTCACATGTCGAGAGCGCGCAAGAGCGCTTCAATGATGTC 529
Db 361 CAGGCCAGCAAGCCCTTTCACATGTCGAGAGCGCGCAAGAGCGCTTCAATGATGTC 420

QY 530 CCTATGAACCGCTGGAATCAAGACCAGTGTCTCAGGGCCCGGAGATGAGCCCAAAATC 589
Db 421 CCTATGAACCGCTGGAATCAAGACCAGTGTCTCAGGGCCCGGAGATGAGCCCAAAATC 480

QY 590 TGCCTGCGCACCCTGCTTCTTCTCTTCTATCAATTAACAGCACTGTGGTGGCCAC 649
Db 481 TGCCTGCGCACCCTGCTTCTTCTCTTCTCAACAATTAACAGCACTGTGGTGGCCAC 540

QY 650 ATGAGAGCAGGCGAGGCGGCGCTGACCTTGGCCACCAAGTTTATCCATATGTCCTG 709
Db 541 ATGAGAGCAGGCGAGGCGGCGCTGACCTTGGCCACCAAGTTTATCCATATGTCCTG 600

QY 710 GATGACCCCAAGTCTGCGGGTGTGGCCACTTGTCTATACAGGAAGATTCGACCGCTTC 769
Db 601 GATGACCCCAAGTCTGCGGGTGTGGCCACTTGTCTATACAGGAAGATTCGACCGCTTC 660

QY 770 ACTGGTACTGTGTGCCCCACAGCCTCTGGGAAGTTTCAGAGGCGCTCAGAGGCTG 829
Db 661 ACTGGTACTGTGTGCCCCACAGCCTCTGGGAAGTTTCAGAGGCGCTCAGAGGCTG 720

QY 830 CGAATCCTGATGAGGAAGTGCATGAGTCCGAGTGGAGTCAATCACTCCCTCTCTCT 889
Db 721 CGAATCCTGATGAGGAAGTGCATGAGTCCGAGTGGAGTCAATCACTCCCTCTCTCT 780

QY 890 GCGCTGAAAGAAAGAGAGAGATGATCTGATCGGTACCCAGAGAGGAGCAAGAAATCC 949
Db 781 GCGCTGAAAGAAAGAGAGAGATGATCTGATCGGTACCCAGAGAGGAGCAAGAAATCC 840

QY 950 AAGATTGCTTGAACCTGCTGAGTTTCAGACTGACAGCCAGGGAAGATGCTCTGAGC 1009
Db 841 AAGATTGCTTGAACCTGCTGAGTTTCAGACTGACAGCCAGGGAAGATGCTCTGAGC 900

QY 1010 CAGGAGAAAGAGAGTGTGTGAGCCCTTCAGTCTGCTTCCGGAAGTGAATTCATGCGC 1069
Db 901 CAGGAGAAAGAGAGTGTGTGAGCCCTTCAGTCTGCTTCCGGAAGTGAATTCATGCGC 960

QY 1070 AAGGCGGAGTGAAGCGGAGTGAATAGCTTGGGCGCATGTTCTGAGAGCGGCGCCAG 1129
Db 961 AAGGCGGAGTGAAGCGGAGTGAATAGCTTGGGCGCATGTTCTGAGAGCGGCGCCAG 1020

QY 1130 CAGTGCCTCCAGCTCGTCTCTCCCGCGGCGCTGTTCAATCCGAGACAGAGAAATGAG 1189
Db 1021 CAGTGCCTCCAGCTCGTCTCTCCCGCGGCGCTGTTCAATCCGAGACAGAGAAATGAG 1080

QY 1190 GAGCAGCGCTACGCTCTGCAAGGCTGCCAGGAATGTCAGGCGTATGAGGTATC 1249
Db 1081 GAGCAGCGCTACGCTCTGCAAGGCTGTCCAGGAATGTCAGGCGTATGAGGTATC 1140

QY 1250 GAGGAGTCCACCAAGCTGTGATCAATGTTCAATGACATCTTCTATCCCTTCCCAATCA 1309
Db 1141 GAGGAGTCCACCAAGCTGTGATCAATGTTCAATGACATCTTCTATCCCTTCCCAATCA 1200

QY 1310 GAGGAGAGAGAGAGTCTGCTCTTCTCCGCGCAATGATGCAAGACCGGCTTCTGCAAT 1369
Db 1201 GAGGAGAGAGAGAGTCTGCTCTTCTCCGCGCAATGATGCAAGACCGGCTTCTGCAAT 1260

QY 1370 TTGTACAAAGTCAACCGCGTTTAAATCCAGAGGCTAGCATGAGTGAAGCCCTCAGC 1429
Db 1261 TTGTACAAAGTCAACCGCGTTTAAATCCAGAGGCTAGCATGAGTGAAGCCCTCAGC 1320

QY 1430 CCCGGGAAGATGAATTTAAGTCCCATTTAAGAAAGAGATGCTCTGACCAAGCGTGA 1489
Db 1321 CCCGGGAAGATGAATTTAAGTCCCATTTAAGAAAGAGATGCTCTGACCAAGCGTGA 1380

QY 1490 TGGGAGTTTGGCGAGGCAAGGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTG 1549
Db 1381 TGGGAGTTTGGCGAGGCAAGGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTG 1440

QY 1550 TACTTCCAGGGACCAAGGACAGCGGCTGTGAGACACCTCTACGTGTGTCACTATGAG 1609
Db 1441 TACTTCCAGGGACCAAGGACAGCGGCTGTGAGACACCTCTACGTGTGTCACTATGAG 1500

QY 1610 GCGGCGGCGAGATGATGATGCTCACACAGCGGCTTCCCATAGCTGTCCATAGAG 1669
Db 1501 GCGGCGGCGAGATGATGATGCTCACACAGCGGCTTCCCATAGCTGTCCATAGAG 1560

QY 1670 CAGAACTTGACATGTTCTGTACAGCACTACAGAGCGGTGAACAGCGGCGCTGTGAC 1729
Db 1561 CAGAACTTGACATGTTCTGTACAGCACTACAGAGCGGTGAACAGCGGCGCTGTGAC 1620

QY 1730 GTCTACAGGTGAGCGGCGCGGAGAGAGCCCTGTACAGAGAGCGGCGCTGTGGCT 1789
Db 1621 GTCTACAGGTGAGCGGCGCGGAGAGAGCCCTGTACAGAGAGCGGCGCTGTGGCT 1680

QY 1790 AGCATATGAGAGAGAGAGAGTGTGCTGAGAGTATGTTCTTCCAGAGATCTTCAATTC 1849
Db 1681 AGCATATGAGAGAGAGAGAGTGTGCTGAGAGTATGTTCTTCCAGAGATCTTCAATTC 1740

QY 1850 CACAGCGCTCGGATGTGGGCTCTACGGCATGATCTACAGCCCAAGCGCTTGACGCA 1909
Db 1741 CACAGCGCTCGGATGTGGGCTCTACGGCATGATCTACAGCCCAAGCGCTTGACGCA 1800

QY 1910 GGGAGAGACCCCAAGCGCTTGTATGAGAGGCGCCAGGTGAGAGCTGGAGAT 1969
Db 1801 GGGAGAGACCCCAAGCGCTTGTATGAGAGGCGCCAGGTGAGAGCTGGAGAT 1860

QY 1970 AACTCTTCAAAAGCATCAAGTACTTGGGCTCAACACACTGGGCTCTCCCTGGGCTAGCC 2029
Db 1861 AACTCTTCAAAAGCATCAAGTACTTGGGCTCAACACACTGGGCTCTCCCTGGGCTAGCC 1920

QY 2030 GTGGTTGTATGAGCGGAGGAGTGTCTCTGACGAGAGGCTTGTGTAAGGAGGCGCTG 2089
Db 1921 GTGGTTGTATGAGCGGAGGAGTGTCTCTGACGAGAGGCTTGTGTAAGGAGGCGCTG 1980

QY 2090 AAAAACAATAGGCGCAGGTGAGATGAGAGACAGTGAAGGCGCTGAGATTCGTGCGC 2149
Db 1981 AAAAACAATAGGCGCAGGTGAGATGAGAGACAGTGAAGGCGCTGAGATTCGTGCGC 2040

QY 2150 GAGAATATAGCTTCAATCAACCTGAGCGAGCTGATCCATCAAGCTGTGCTTCAAGGAGG 2209
Db 2041 GAGAATATAGCTTCAATCAACCTGAGCGAGCTGATCCATCAAGCTGTGCTTCAAGGAGG 2100

QY 2210 TTCTCTGCTCATGAGGAGCTAATCCAAAGCCCAAGTGTTCAGAGTGTGACATCGCGGCT 2269

Db 2101 TTCTCTCGCTCAATGGGGCTAATCCACAGCCAGGCTGTTCAAGGATGGCCATGCGGGT 2160
QY 2270 GCCCGGCTACCGTCTGTGGATGGCTTACGACACAGGGTACTGAGCCCTACATGAGCGTC 2329
Db 2161 GCGCCGGTCCACCGTCTGTGGATGGCTTACGACACAGGGTACTGAGCCCTACATGAGCGTC 2220
QY 2230 CCTGAGAACACACAGCAGCGGCTATGAGGGGGTTCGTGGGCCCTGACACGTGGAGAACCTG 2389
Db 2221 CCTGAGAACACACAGCAGCGGCTATGAGGGGGTTCGTGGGCCCTGACACGTGGAGAACCTG 2280
QY 2290 CCCAATGAGCCCAACCGCTTGTATCTCCACAGGCTTCTGACGAAACGTCACATTT 2449
Db 2281 CCCAATGAGCCCAACCGCTTGTATCTCCACAGGCTTCTGACGAAACGTCACATTT 2340
QY 2450 TTCCACACAAACTTCTGCTGTCTCCCAACTGATTCGACAGGAGAAACCTTACACCTCAG 2509
Db 2341 TTCCACACAAACTTCTGCTGTCTCCCAACTGATTCGACAGGAGAAACCTTACACCTCAG 2400
QY 2510 ATCTACCCCAACGAGACACAGTATCGCTGGCCCGAGTGGGCGAGCACTATGAAGTC 2569
Db 2401 ATCTACCCCAACGAGACACAGTATCGCTGGCCCGAGTGGGCGAGCACTATGAAGTC 2460
QY 2570 ACCTTGGCTGCACTTTCTACAGGAATACCTCTGAGC 2604
Db 2461 ACCTTGGCTGCACTTTCTACAGGAATACCTCTGAGC 2495

RESULT 8
ABK83337
ID ABK83337 standard; cDNA; 4076 BP.
XX
AC ABK83337;
XX
DT 12-AUG-2002 (first entry)
XX
DE cDNA encoding human DPRP-2 splice variant #5.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Q1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI: 2002-444178/47.
XX
DR P-PDB; ABG61606.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT -
XX
PS Disclosure; Page 90-91; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)

CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83322-ABK83343 encode human DPRP proteins.
XX

SO Sequence 4076 BP; 879 A; 1276 C; 1143 G; 778 T; 0 other;

Query Match 87.7%; Score 2296.4; DB 24; Length 4076;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 2450; Conservative 0; Mismatches 1; Indels 143; Gaps 1;

QY 11 ATGGCCACACCGGGACCCCAAGCGCGAGCGAGCGAGCGAGCGCGACAGATGACCG 70
Db 367 ATGGCCACACCGGGACCCCAAGCGCGAGCGAGCGAGCGAGCGCGACAGATGACCG 426
QY 71 GCGCCCGCTTCCAGGTGACAGAACTCGTGGAGCGGGCTCCGAGCATATCATCAGGC 130
Db 427 GCGCCCGCTTCCAGGTGACAGAACTCGTGGAGCGGGCTCCGAGCATATCATCAGGC 486
QY 131 AGCCGCAAGTACTGGGCTCATTTGTCAACAGAGCGCCCAAGCACTTCCAGTTTGTGAG 190
Db 487 AGCCGCAAGTACTGGGCTCATTTGTCAACAGAGCGCCCAAGCACTTCCAGTTTGTGAG 546
QY 191 AAGAGGATGATGTGGGGCCCACTCCACAGCGCTACTACTACTCGGGGAATGCCATATGAGC 250
Db 547 AAGAGGATGATGTGGGGCCCACTCCACAGCGCTACTACTACTCGGGGAATGCCATATGAGC 606
QY 251 AGCCGCAAGTACTGGGCTCATTTGTCAACAGAGCGCGCGAGCGCGAGCGCGAGCGCG 310
Db 607 AGCCGCAAGTACTGGGCTCATTTGTCAACAGAGCGCGCGAGCGCGAGCGCGAGCGCG 666
QY 311 CTGCTCTCTCTCTGGAAGCAGATGCTGATATTTTCCAGGCGCCACCGCATATGGGTC 370
Db 667 CTGCTCTCTCTCTGGAAGCAGATGCTGATATTTTCCAGGCGCCACCGCATATGGGTC 726
QY 371 TACTCTGGGGGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 430
Db 727 TACTCTGGGGGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 786
QY 431 TCTTACGACTTCCACAGCAGAGAGTGGCTTCTCTCTCTCCAGGCGCAGCAAGCCTCTTC 490
Db 787 TCTTACGACTTCCACAGCAGAGAGTGGCTTCTCTCTCTCCAGGCGCAGCAAGCCTCTTC 846
QY 491 CACTGTGCGCAGCGCGCAGAGAGCGCTTCAATGCTGCTCCCTATGAACCGCTGGAATC 550
Db 847 CACTGTGCGCAGCGCGCAGAGAGCGCTTCAATGCTGCTCCCTATGAACCGCTGGAATC 906
QY 551 AAGACCCAGTGTCCAGGGCCCCGGATGAGACCCCAAAATCTGCGCGGACCGCTGCTTC 610
Db 907 AAGACCCAGTGTCCAGGGCCCCGGATGAGACCCCAAAATCTGCGCGGACCGCTGCTTC 966
QY 611 TTCTCTTCAATCAATAACAGGACCTGTGGGTGGCCAAATCGAGACAGCGGAGAGCGG 670
Db 967 TTCTCTTCAATCAATAACAGGACCTGTGGGTGGCCAAATCGAGACAGCGGAGAGCGG 1026
QY 671 CGGCTGACCTTCTGCGCACCAAGTTTATCAATGCTCTGGATGACCCCACTGCGGGT 730
Db 1027 CGGCTGACCTTCTGCGCACCAAGTTTATCCATGCTCTGGATGACCCCACTGCGGGT 1086
QY 731 GTGGCCACCTTCTGATACAGAGAGTTCAGACCGCTTCACTGGGTACTGTGGTGCC 790
Db 1087 GTGGCCACCTTCTGATACAGAGAGTTCAGACCGCTTCACTGGGTACTGTGGTGCC 1146
QY 791 ACAGCTCTCTTGGAGAGTTTCAGAGGCGCTCAAGACGCTCGCAATCTCTATGAGAAATC 850
Db 1147 ACAGCTCTCTTGGAGAGTTTCAGAGGCGCTCAAGACGCTCGCAATCTCTATGAGAAATC 1206

OY 851 GATGAGTCGAGGTGAGGTGATTCATCTCCCTCTCTGCTAGAGAAAGAGACG 910
 |||||||
 Db 1207 GATGAGTCGAGGTGAGGTGATTCATCTCCCTCTCTGCTAGAGAAAGAGACG 1266
 OY 911 GACTCGTATCGTATACCCAGAGGACGAGCAAGATCCCAAGTTGGCTTTGAAGTGGCT 970
 |||||||
 Db 1267 GACTCGTATCGTATACCCAGAGGACGAGCAAGATCCCAAGTTGGCTTTGAAGTGGCT 1326
 OY 971 GAGTTCAGACTGACAGCCAGGCGCAAGATGCTCGACCCAGAGAGAGAGTGGTGCAG 1030
 |||||||
 Db 1327 GAGTTCAGACTGACAGCCAGGCGCAAGATGCTCGACCCAGAGAGAGAGTGGTGGCAG 1386
 OY 1031 CCCCTTACGCTGCTGTTCCCGAAGGTGAGATACATCGCCAGGCGCGGTGGAGCCGGGAT 1090
 |||||||
 Db 1387 CCTTCAGCTGCTGTTCCCGAAGGTGAGATACATCGCCAGGCGCGGTGGAGCCGGGAT 1446
 OY 1091 GCGAATACGCTGAGGCGCATGTTCTGAGACGCGGCCAGAGATGGTCCAGCTGCTGC 1150
 |||||||
 Db 1447 GCGAATACGCTGAGGCGCATGTTCTGAGACGCGGCCAGAGATGGTCCAGCTGCTGC 1506
 OY 1151 CTCCTCCCGGCTGTTTCATCCGAGCACAGAAATAGAGAGCGGCTAGGCTCTGCGC 1210
 |||||||
 Db 1507 CTCCTCCCGGCTGTTTCATCCGAGCACAGAAATAGAGAGCGGCTAGGCTCTGCGC 1566
 OY 1211 AGAGCTGTCGCCAGGAATGTCCAGCCGTATGTGTACGAGGAGGTACCAACGTCTGG 1270
 |||||||
 Db 1567 AGAGCTGTCGCCAGGAATGTCCAGCCGTATGTGTACGAGGAGGTACCAACGTCTGG 1626
 OY 1271 ATCAATGTTCAATACATCTTCTATCCCTTCCCATCATGAGAGGAGAGAGAGCTCTGC 1330
 |||||||
 Db 1627 ATCAATGTTCAATACATCTTCTATCCCTTCCCATCATGAGAGGAGAGAGAGCTCTGC 1686
 OY 1331 TTTCTCCGCGCCAAATATGACAGACCGGCTTCTGCCATTTGTACAAAGTACACGCGGT 1390
 |||||||
 Db 1687 TTTCTCCGCGCCAAATATGACAGACCGGCTTCTGCCATTTGTACAAAGTACACGCGGT 1746
 OY 1391 TTTAAATCCAGGGCTACGATTTGAGTGAAGCCCTTCAAGCCCGGGGAAGATATTTAAG 1450
 |||||||
 Db 1747 TTTAAATCCAGGGCTACGATTTGAGTGAAGCCCTTCAAGCCCGGGGAAGATATTTAAG 1806
 OY 1451 TGGCCCATTAAGGAAGATGCTGTACACAGGGGTGAATGGAGGTTTGGGAGGAC 1510
 |||||||
 Db 1807 TGGCCCATTAAGGAAGATGCTGTACACAGGGGTGAATGGAGGTTTGGGAGGAC 1866
 OY 1511 GGGTCCAAAGTCTGGGTCAATGAGAGACCAAGCTGGTACTTCCAGGGCACCAAGAC 1570
 |||||||
 Db 1867 GGGTCCAAAGTCTGGGTCAATGAGAGACCAAGCTGGTACTTCCAGGGCACCAAGAC 1926
 OY 1571 AGCGCGCTGAGAGACCACTCTACGTGTAGCTATGAGGCGCGCGAGATCTACGC 1630
 |||||||
 Db 1927 AGCGCGCTGAGAGACCACTCTACGTGTAGCTATGAGGCGCGCGAGATCTACGC 1986
 OY 1631 CTACACAGCGCGGCTTCTCCCTACGTGTCTCATGAGCGAGAACTTCGATGTTGCTC 1690
 |||||||
 Db 1987 CTACACAGCGCGGCTTCTCCCTACGTGTCTCATGAGCGAGAACTTCGATGTTGCTC 2046
 OY 1691 AGCCACTAGACAGCGTGAAGCACGCGCTCGGTGACAGTCTACAAAGCTGAGCGGCC 1750
 |||||||
 Db 2047 AGCCACTAGACAGCGTGAAGCACGCGCTCGGTGACAGTCTACAAAGCTGAGCGGCC 2106
 OY 1751 GACGAGACCGCTGACAGACAGAGCGCGCTTGGGCTAGCATGTATGAGAGCGACG 1810
 |||||||
 Db 2107 GACGAGACCGCTGACAGACAGAGCGCGCTTGGGCTAGCATGTATGAGAGCGACG 2166
 OY 1811 TGGCCCGCGGATTAATGCTCTCCAGAGATCTCATTTCCACAGCGCGGTGGATGGCG 1870
 |||||||
 Db 2167 TGGCCCGCGGATTAATGCTCTCCAGAGATCTCATTTCCACAGCGCGGTGGATGGCG 2226
 OY 1871 CTCTACGCGATGATCTACAAGCCCGACGCTTTCAGAGCGAGGAAAGACACCCACGCTC 1930
 |||||||
 Db 2227 CTCTACGCGATGATCTACAAGCCCGACGCTTTCAGAGCGAGGAAAGACACCCACGCTC 2286

OY 1931 CTCTTTGATATAGAGAGCCCGCAGGTGACGTGTGATTAATCTTCAAGGCATCAAG 1990
 |||||||
 Db 2287 CTCTTTGATATAGAGAGCCCGCAGGTGACGTGTGATTAATCTTCAAGGCATCAAG 2346
 OY 1991 TACTTGGGCTCAACACACTGGCTCCCTGGGCTAGCGCGTGGTGTGATTTGACGCGAG 2050
 |||||||
 Db 2347 TACTTGGGCTCAACACACTGGCTCCCTGGGCTAGCGCGTGGTGTGATTTGACGCGAG 2406
 OY 2051 GGCCTGTGACGAGGCGTTCGTTTCGAAGGGGCCCTGAAAAACCAATGGAGCGAGTG 2110
 |||||||
 Db 2407 GGCCTGTGACGAGGCGTTCGTTTCGAAGGGGCCCTGAAAAACCAATGGAGCGAGTG 2466
 OY 2111 GAGATGAGAGACAGGTGAGAGGCGCTGCAAGTTCGTGGCCGAGAGATGAGCTTCATGAC 2170
 |||||||
 Db 2467 GAGATGAGAGACAGGTGAGAGGCGCTGCAAGTTCGTGGCCGAGAGATGAGCTTCATGAC 2526
 OY 2171 CTGAGCGAGTTCATCCATGCTGTGCTACGAGGGGCTTCTCTGCTCATGGGGCTA 2230
 |||||||
 Db 2527 CTGAGCGAGTTCATCCATGCTGTGCTACGAGGGGCTTCTCTGCTCATGGGGCTA 2586
 OY 2231 ATCCACAAGCCCGAGGTTCATGAGTGGCCATGCGGGGTGCCCGGTCAACGCTGAGTG 2290
 |||||||
 Db 2587 ATCCACAAGCCCGAGGTTCATGAGTGGCCATGCGGGGTGCCCGGTCAACGCTGAGTG 2610
 OY 2291 GCCTACGACAGAGGTACACTGAGCGCTACATGAGAGCTCCCTGAGAACACAGACGCGC 2350
 |||||||
 Db 2611 GCCTACGACAGAGGTACACTGAGCGCTACATGAGAGCTCCCTGAGAACACAGACGCGC 2610
 OY 2351 TATGAGCGGGTTCGTTGCGCCCTGCAAGTGGAGAACCTGCCAATGAGCCCAACGCGTTG 2410
 |||||||
 Db 2611 TATGAGCGGGTTCGTTGCGCCCTGCAAGTGGAGAACCTGCCAATGAGCCCAACGCGTTG 2623
 OY 2411 CTATCTCCACAGGCTTCTCTGAGAGAAAGCTGTTTCCACAAACTTCTCTGCTC 2470
 |||||||
 Db 2624 CTATCTCCACAGGCTTCTCTGAGAGAAAGCTGTTTCCACAAACTTCTCTGCTC 2683
 OY 2471 TCCCAACTGATCCGAGCAGAGGAAACCTTACAGCTCCAGATCTACCCCAAGAGAGAC 2530
 |||||||
 Db 2684 TCCCAACTGATCCGAGCAGAGGAAACCTTACAGCTCCAGATCTACCCCAAGAGAGAC 2743
 OY 2531 AGTATTCGCTGCGCCGAGTGCAGGCGGAGCACTATGAAGTCACGTTGCTGCACTTACAG 2590
 |||||||
 Db 2744 AGTATTCGCTGCGCCGAGTGCAGGCGGAGCACTATGAAGTCACGTTGCTGCACTTACAG 2803
 OY 2591 GAATACCTCTGAGC 2604
 |||||||
 Db 2804 GAATACCTCTGAGC 2817
 |||||||
 RESULT 9
 ABR83336
 ID ABR83336 standard; cDNA: 4159 BP.
 XX
 AC ABR83336;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE cDNA encoding human DRRP-2 splice variant #4.
 XX
 DE
 XX Humun; serine protease; dipeptidyl peptidase IV-related protein; DRRP;
 KW DRRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyslexia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX

PF 12-OCT-2001: 2001WO-US31874.
 XX
 PR 12-OCT-2000: 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Q1 S, Akinsanya KO, Riviére PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 XX P-PSDB: AB661605.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 PS Disclosure; Page 87-88: 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bullimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABK8332-ABK83343 encode human DPRP proteins.
 XX
 SQ Sequence 4159 BP; 894 A; 1306 C; 1174 G; 785 T; 0 other:

Query Match 87.7%; Score 2296.4; DB 24; Length 4159;
 Best Local Similarity 94.4%; Pred. No. 0;
 Matches 2450; Conservative 0; Mismatches 1; Indels 143; Gaps 1;

QY 11 ATGGCCACACCGGAGCCGACCGGACGAGCGACGCGCCACAGATGACCG 70
 DB ATGGCCACACCGGAGCCGACCGGACGAGCGAGCGCGCCACAGATGACCG 426
 QY 71 GCGCGCGGCTCCAGGAGCAAGACACGCGGAGCGGCTCGGAGCATGACGCG 130
 DB GCGCGCGGCTCCAGGAGCAAGACACGCGGAGCGGCTCGGAGCATGACGCG 486
 QY 131 AGCGCAAGTACTGCGGCTCATTTGTCAACAAAGGCGCCACGATTCAGTTGTGCG 190
 DB AGCGCAAGTACTGCGGCTCATTTGTCAACAAAGGCGCCACGATTCAGTTGTGCG 546
 QY 191 AAGACGATGAGTGGGCGCCACTCCACGCGCTTACTACTGCGGAATCCATATGCG 250
 DB AAGACGATGAGTGGGCGCCACTCCACGCGCTTACTACTGCGGAATCCATATGCG 606
 QY 251 AGCGGAGGAGACTCCCTCTTACTGAGATTCACAAAGAGTCCGAAAGAGGCTGTG 310
 DB AGCGGAGGAGACTCCCTCTTACTGAGATTCACAAAGAGTCCGAAAGAGGCTGTG 666
 QY 311 CTGCTCTGTCCTGGAAGCAGATGCTGATCATTTCCAGGCGCCACCATGAGGCTC 370
 DB CTGCTCTGTCCTGGAAGCAGATGCTGATCATTTCCAGGCGCCACCATGAGGCTC 726
 QY 371 TACTCTGGGAGGAGAGCTCTGAGGAGCGGAAAGCGCTGGGGTCTTGGCATCCG 430
 DB TACTCTGGGAGGAGAGCTCTGAGGAGCGGAAAGCGCTGGGGTCTTGGCATCCG 786
 QY 431 TCTACGAGCTTCCACAGGAGAGTGGCTTCTCTCCAGGCGCAGCAAGAGCTCTTC 490
 DB TCTACGAGCTTCCACAGGAGAGTGGCTTCTCTCTCCAGGCGCAGCAAGAGCTCTTC 846
 QY 491 CACTGTGCGAGCGGCGCAAGAGGCTTCAATGATGTCCTTATGAACCGCTGGAATC 550
 DB CACTGTGCGAGCGGCGCAAGAGGCTTCAATGATGTCCTTATGAACCGCTGGAATC

DB 847 CACTGTGCGAGCGGCGCAAGAGGCTTCAATGATGTCCTTATGAACCGCTGGAATC 906
 QY 551 AAGACCGAGTGTGAGGGGCGCGGATGAGCCCAAAATGTGCGCGAGCCGCTTC 610
 DB AAGACCGAGTGTGAGGGGCGCGGATGAGCCCAAAATGTGCGCGAGCCGCTTC 966
 QY 611 TTCTCTTCAATCAATAACAGCGACCTGTGGTGCCCAATCAGACGCGAGGACGG 670
 DB TTCTCTTCAATCAATAACAGCGACCTGTGGTGCCCAATCAGACGCGAGGACGG 1026
 QY 671 CGGCTGACCTTGTGCGACCAAGGTTATCCATATGCTCGATGACCCCAAGTCCGGGT 730
 DB CGGCTGACCTTGTGCGACCAAGGTTATCCATATGCTCGATGACCCCAAGTCCGGGT 1086
 QY 731 GTGGCGACCTTGTGCAATACAGGAAGTGTGACCGCTTCACTGGGTCTGAGTGGTCCG 790
 DB GTGGCGACCTTGTGCAATACAGGAAGTGTGACCGCTTCACTGGGTCTGAGTGGTCCG 1146
 QY 791 ACACCTCTGTTGGGAGGTTCCAGAGGCGCTCAAGACGCTGCAATCTGTATGAGGAATC 850
 DB ACACCTCTGTTGGGAGGTTCCAGAGGCGCTCAAGACGCTGCAATCTGTATGAGGAATC 1206
 QY 851 GATGAGTCCGAGGTGAGGTCATTCACGTCCTCTCTGCGCTAGAAAGAAAGACG 910
 DB GATGAGTCCGAGGTGAGGTCATTCACGTCCTCTCTGCGCTAGAAAGAAAGACG 1266
 QY 911 GACTCGATGCTGATCCCGACGACGACGCAAGATCCCAATGCTGCTTAAACTGGCT 970
 DB GACTCGATGCTGATCCCGACGACGACGCAAGATCCCAATGCTGCTTAAACTGGCT 1336
 QY 971 GAGTTCCAGATGACAGCCAGGAGCAAGTCTCTGACCCAGAGAAAGAGAGCTGTGCG 1030
 DB GAGTTCCAGATGACAGCCAGGAGCAAGTCTCTGACCCAGAGAAAGAGAGCTGTGCG 1386
 QY 1327 GAGTTCCAGATGACAGCCAGGAGCAAGTCTCTGACCCAGAGAAAGAGAGCTGTGCG 1386
 DB GAGTTCCAGATGACAGCCAGGAGCAAGTCTCTGACCCAGAGAAAGAGAGCTGTGCG 1446
 QY 1447 GGCATAATGAGCTGTGGCCATTTCTTGAGCCGCGCCAGCGTGCACCTGCTCTC 1506
 DB GGCATAATGAGCTGTGGCCATTTCTTGAGCCGCGCCAGCGTGCACCTGCTCTC 1550
 QY 1511 CTCCCGCGGCTTTCATCCGAGCAGAGATGAGAGAGCGGCTGACCTTCC 1210
 DB CTCCCGCGGCTTTCATCCGAGCAGAGATGAGAGAGCGGCTGACCTTCC 1566
 QY 1507 CTCCCGCGGCTTTCATCCGAGCAGAGATGAGAGAGCGGCTGACCTTCC 1566
 DB CTCCCGCGGCTTTCATCCGAGCAGAGATGAGAGAGCGGCTGACCTTCC 1626
 QY 1211 AGAGCTGTCCAGGAATGTCCAGCCGTATGTGTGTAAGAGAGTCCACACGCTGG 1270
 DB AGAGCTGTCCAGGAATGTCCAGCCGTATGTGTGTAAGAGAGTCCACACGCTGG 1626
 QY 1271 ATCAATGTTCATGACATCTTATCCCTCCCAATCAGAGAGGAGAGAGAGCTGTGC 1330
 DB ATCAATGTTCATGACATCTTATCCCTCCCAATCAGAGAGGAGAGAGAGCTGTGC 1686
 QY 1627 ATCAATGTTCATGACATCTTATCCCTCCCAATCAGAGAGGAGAGAGAGCTGTGC 1330
 DB ATCAATGTTCATGACATCTTATCCCTCCCAATCAGAGAGGAGAGAGAGCTGTGC 1686
 QY 1331 TTCTCTCCGCGCAATGATGACAGCGGCTTGTGCAATTTGTCAAAAGTCCAGCGCTT 1390
 DB TTCTCTCCGCGCAATGATGACAGCGGCTTGTGCAATTTGTCAAAAGTCCAGCGCTT 1746
 QY 1687 TTCTCTCCGCGCAATGATGACAGCGGCTTGTGCAATTTGTCAAAAGTCCAGCGCTT 1390
 DB TTCTCTCCGCGCAATGATGACAGCGGCTTGTGCAATTTGTCAAAAGTCCAGCGCTT 1746
 QY 1391 TTAAATCCAGGCTTACGATGTGAGTACGCTTCCAGCCCGGGGGAAGATTAATTAAG 1450
 DB TTAAATCCAGGCTTACGATGTGAGTACGCTTCCAGCCCGGGGGAAGATTAATTAAG 1806
 QY 1451 TGCCCATTAAGGAAGATGCTGTGACAGCGGCTGAAATGGAGGTTTGGCGAGGAC 1510
 DB TGCCCATTAAGGAAGATGCTGTGACAGCGGCTGAAATGGAGGTTTGGCGAGGAC 1866
 QY 1807 TGCCCATTAAGGAAGATGCTGTGACAGCGGCTGAAATGGAGGTTTGGCGAGGAC 1510
 DB TGCCCATTAAGGAAGATGCTGTGACAGCGGCTGAAATGGAGGTTTGGCGAGGAC 1866
 QY 1511 GGTCTCAAGATCTGGGTCAATGAGAGACCAAGGCTGTGTAATCTTCCAGGCGCAAGGAC 1570
 DB GGTCTCAAGATCTGGGTCAATGAGAGACCAAGGCTGTGTAATCTTCCAGGCGCAAGGAC 1926
 QY 1867 GGTCTCAAGATCTGGGTCAATGAGAGACCAAGGCTGTGTAATCTTCCAGGCGCAAGGAC 1570
 DB GGTCTCAAGATCTGGGTCAATGAGAGACCAAGGCTGTGTAATCTTCCAGGCGCAAGGAC 1926
 QY 1571 AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGCGCGCGAGATGCTACGC 1630
 DB AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGCGCGCGAGATGCTACGC 1986
 QY 1927 AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGCGCGCGAGATGCTACGC 1630
 DB AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGCGCGCGAGATGCTACGC 1986

Oy	1631	CTACACAGCCCGGGCTCTCCCAATAGCTGCTCCATAGACACAACTGTGACATGTTGCTC	1630
Db	1987	CTCACACAGCCCGGGCTCTCCCAATAGCTGCTCCATAGACACAACTGTGACATGTTGCTC	2046
Oy	1691	AGCCACTACAGACACGTCGACACAGCCGCCCTCGGTCACAGTGTACAAAGCTGAGCGGCC	1750
Db	2047	AGCCACTACAGACACGTCGACACAGCCGCCCTCGGTCACAGTGTACAAAGCTGAGCGGCC	2106
Oy	1751	GACGACGACCCCTGACACAGACAGCCCGCTTCTG6GGCTAGCATGTGAGGACGCCAGC	1810
Db	2107	GACGACGACCCCTGACACAGACAGCCCGCTTCTG6GGCTAGCATGTGAGGACGCCAGC	2166
Oy	1811	TGCCCCCGGGATTATGTCTCTCCAGAGATCTTCATTCTCACACCGCGCTCGGATGTGGG	1870
Db	2167	TGCCCCCGGGATTATGTCTCTCCAGAGATCTTCATTCTCACACCGCGCTCGGATGTGGG	2226
Oy	1871	CTTACGGCATGATCTCAAGCCCCCAGCGCTTGACAGCCAGGGAGAAGACGCCACGCTC	1930
Db	2227	CTTACGGCATGATCTCAAGCCCCCAGCGCTTGACAGCCAGGGAGAAGACGCCACGCTC	2286
Oy	1931	CTCTTTGTATATGAGAGCCGCCAGGTGACGTGTGTGAATACTCTTCAAAAGCATCAAG	1990
Db	2287	CTCTTTGTATATGAGAGCCGCCAGGTGACGTGTGTGAATACTCTTCAAAAGCATCAAG	2346
Oy	1991	TACTTGGGGCTCAACACACTGAGGCTCCGAGGGCTACGCGTG6TGTGATATGAGCGCAG	2050
Db	2347	TACTTGGGGCTCAACACACTGAGGCTCCGAGGGCTACGCGTG6TGTGATATGAGCGCAG	2406
Oy	2051	GGCTCTGTGTAGCAGGAGGGCTTTCGGTTTCGAAGGGGGCCCTTAAAAACCAATG6GCCAGTG	2110
Db	2407	GGCTCTGTGTAGCAGGAGGGCTTTCGGTTTCGAAGGGGGCCCTTAAAAACCAATG6GCCAGTG	2466
Oy	2111	GAGATCCAGGACACAGGTGTGAGGGGCTGTGCAATTCGTGTG6CGAAGATATG6CTTCA TGAC	2170
Db	2467	GAGATCCAGGACACAGGTGTGAGGGGCTGTGCAATTCGTGTG6CGAAGATATG6CTTCA TGAC	2526
Oy	2171	CTGAGCCGAGTGTGCAATTCAGTGTG6CTTACAGGGGGCTTCTGTGCTCATG6GGCTA	2230
Db	2527	CTGAGCCGAGTGTGCAATTCAGTGTG6CTTACAGGGGGCTTCTGTGCTCATG6GGGCTA	2586
Oy	2231	ATCCACAAGCCCCAGGTGTTCAAAGTGTG6CCATTCGGGGGTCCCGGCTACCGTGTGATG	2290
Db	2587	ATCCACAAGCCCCAGGTGTTCAAAGTGTG6CCATTCGGGGGTCCCGGCTACCGTGTGATG	2610
Oy	2291	GCTACAGACACAGG6GTACACTGAGGCGCTACATGTGAGCTGCCGTGAGAACACAGACGCGC	2350
Db	2611	GCTACAGACACAGG6GTACACTGAGGCGCTACATGTGAGCTGCCGTGAGAACACAGACGCGC	2610
Oy	2351	TATGAGGCGGGTTCCGTG6GCTGTGACAGTGTGAGAAAGCTGCCAATGAGCCCAACGCGTTG	2410
Db	2611	TATGAGGCGGGTTCCGTG6GCTGTGACAGTGTGAGAAAGCTGCCAATGAGCCCAACGCGTTG	2623
Oy	2411	CTTATCTCCACAGGCTCTCTGTGAGCAAGAAAAGTGTACACTTTTTCACACAAACTTCTCTGTC	2470
Db	2624	CTTATCTCCACAGGCTCTCTGTGAGCAAGAAAAGTGTACACTTTTTCACACAAACTTCTCTGTC	2683
Oy	2471	TCCCAACTGATGTCGAGAGGGAAGAACTTCCACAGCTCCAAATCTTACCCCAAGAGAGAC	2530
Db	2684	TCCCAACTGATGTCGAGAGGGAAGAACTTCCACAGCTCCAAATCTTACCCCAAGAGAGAC	2743
Oy	2531	AGTATTCGTGCCCCGAGTGTGGGCGAGCACTATGAAGTACACTTGTCTGCTACTTCTACAG	2590
Db	2744	AGTATTCGTGCCCCGAGTGTGGGCGAGCACTATGAAGTACACTTGTCTGCTACTTCTACAG	2803
Oy	2591	GAATACTCTGTAGC 2604	
Db	2804	GAATACTCTGTAGC 2817	

XX	AAI57896;
AC	
XX	
DT	22-Oct-2001 (first entry)
XX	
DE	
XX	Human polynucleotide SEQ ID NO 99.
KW	Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Dragar Syndrome; chemotactic;
KW	chemokineic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-Jul-2001.
XX	
XX	26-Dec-2000; 2000WO-US34263.
PF	
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QB, Zhou P, Goodrich R, Drmanac RT;
DR	WPI: 2001-442253/47.
DR	P-PSDB: AAM38740.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
PS	Claim 1; SEQ ID NO 99; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic.
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Dragar Syndrome. Other uses include the
CC	utilistation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemia and
CC	C.N.S. disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 2801 BP; 586 A; 889 C; 801 G; 525 T; 0 other;
Query Match	87.4%; Score 2288.2; DB 22; Length 2801;
Best Local Similarity	94.1%; Pred. No. 0;
Matches 2485; Conservative	0; Mismatches 23; Indels 133; Gaps 5;
OY	11 ATGGCGACCAGGAGGCCCAAGCGCCGACGAGCGCAGCGGCACAGATGACCG 70
Ddb	245 ATGGCGACCAGGAGGCCCAAGCGCCGACGAGCGCAGCGGCACAGATGACCG 304
OY	71 GCCGCCCGCTTCAGAGTGACAGAGCACACTGTTGGAGACGGGCTCCGGAGCATCATCCACGC 130

|||||
Db 305 GCCGCCGCTTCCAGGTCGAGAAGCACTCGTGGAGCGGGCTCCGGAGCATCATCCAGCGC 364
QY 131 AGCGGAGACTCTCGGGCCCTCATTTGCAACAAGCGCCGCCACGACTTCCAGTTGTGACAG 190
Db 365 AGCGGAGAGTACGCGGGCCCTCATTTGCAACAAGCGCCGCCACGACTTCCAGTTGTGACAG 424
QY 191 AAGACGATAGTCTGGGGCCCACTCCACCGCTCTACTACCTGGGGAATGCCATATGGC 250
Db 425 AAGACGATAGTCTGGGGCCCACTCCACCGCTCTACTACCTGGGGAATGCCATATGGC 484
QY 251 AGCGGAGAGACTCCCTCTCTACTCTGAGATTCCCAAGAAGTCCGAAAGAGGCTCTG 310
Db 485 AGCGGAGAGACTCCCTCTCTACTCTGAGATTCCCAAGAAGTCCGAAAGAGGCTCTG 544
QY 311 CTGCTCTCTCTGCGAAGCAGATGCTGGATCATTTCCAGGCGACCGCCACCATGGGCTC 370
Db 545 CTGCTCTCTCTGCGAAGCAGATGCTGGATCATTTCCAGGCGACCGCCACCATGGGCTC 604
QY 371 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAACGCGTGGGGGCTTCCGCTATACC 430
Db 605 TACTCTCGGAGAGAGAGTCTGAGGAGCGGAACGCGTGGGGGCTTCCGCTATACC 664
QY 431 TCCTAGACTTCCACAGCAGAGTGGCTCTCTCTTCCAGGCGACCAACAGCTCTTC 490
Db 665 TCCTAGACTTCCACAGCAGAGTGGCTCTCTCTTCCAGGCGACCAACAGCTCTTC 724
QY 491 CACTCTGCGACGCGCGCAAGAACGCTTCATC-----GTGTCC 529
Db 725 CACTCTGCGACGCGCGCAAGAACGCTTCATGCTGAGCCCTGTGCTGTCTC 784
QY 530 CCTATGAAACCGGTGAAATCAAGAACCCAGTGTCAAGGCGCCGGATGAGCCCAATC 589
Db 785 CCTATGAAACCGGTGAAATCAAGAACCCAGTGTCAAGGCGCCGGATGAGCCCAATC 844
QY 590 TGGCTTCGCGACCCCTGCTCTCTCTCTCTCATCAATTAACAGCGACTGTGGTGGCAAC 649
Db 845 TGGCTTCGCGACCCCTGCTCTCTCTCTCTCATCAATTAACAGCGACTGTGGTGGCAAC 904
QY 650 ATGAGAGACAGCGGAGAGCGCGGCTGACCTTCTGCCACCAAGTTTATCATGTCTGTG 709
Db 905 ATGAGAGACAGCGGAGAGCGCGGCTGACCTTCTGCCACCAAGTTTATCATGTCTGTG 964
QY 710 GATGACCCCAAGTCTGCGGGGTGGGCCACCTGTGCTATACAGGAAGTTCGACCGGCTC 769
Db 965 GATGACCCCAAGTCTGCGGGGTGGGCCACCTGTGCTATACAGGAAGTTCGACCGGCTC 1024
QY 770 ACTGTGACTGTGTGTGTCGCCACAGCTCTGTGGAAAGTTCAAGAGGCTTCAAGACGCTG 829
Db 1025 ACTGTGACTGTGTGTGTCGCCACAGCTCTGTGGAAAGTTCAAGAGGCTTCAAGACGCTG 1084
QY 830 CGAATCTCTGTATGAGGAAGTGTGATGAGTCCGAGGTGGAAGTCAATTCAGCTCCCTCTCT 889
Db 1085 CGAATCTCTGTATGAGGAAGTGTGATGAGTCCGAGGTGGAAGTCAATTCAGCTCCCTCTCT 1144
QY 890 GCGCTAGAAAGAAAGAGAGAGGACTGTATCGGTATCCAGAGAGGACAGCAAGAAATCCC 949
Db 1145 GCGCTAGAAAGAAAGAGAGGACTGTATCGGTATCCAGAGAGGACAGCAAGAAATCCC 1204
QY 950 AAGATTCCTTTGAACCTGGCTGAGTTCCAGATGACAGCCAGGGCAAGATCTGTGAC 1009
Db 1205 AAGATTCCTTTGAACCTGGCTGAGTTCCAGATGACAGCCAGGGCAAGATCTGTGAC 1264
QY 1010 CAGAGAGAGAGAGTGTGTGACCCCTTTCAGTCTGCTTCCGAAGGTGAGTACATGGCC 1069
Db 1265 CAGAGAGAGAGAGTGTGTGACCCCTTTCAGTCTGCTTCCGAAGGTGAGTACATGGCC 1324
QY 1070 AGGGCCGCGGTGAGCCCGGATGGAATAGCCCTGGGCGCATGTCTCGAGACCGGCCAG 1129
Db 1325 AGGGCCGCGGTGAGCCCGGATGGAATAGCCCTGGGCGCATGTCTCGAGACCGGCCAG 1384
QY 1130 CAGTGTCTCAGCTGTCTCTCTCCCGCGCCCTGTTCATCCGAGACAGAAATGAG 1189
|||||

Db 1385 CAGTGTCTCAGCTGTCTCTCTCCCGCGCCCTGTTCAATCCGAGACAGAAATGAG 1444
QY 1190 GAGCA--GCGGCTAGCCCTCTGCGAGAGCTGTCCCGAAGATGTCCAGCGATGTGGTGA 1248
Db 1445 GAGCAGCGGCTAGCCCTCTGCGAGAGCTGTCCCGAAGATGTCCAGCGATGTGGTGA 1504
QY 1249 CGAGAGAGTACCAACGCTGTGATCAATGTTTCATGATCATCTTCTGATCTCTCCCAATC 1308
Db 1505 GAGAGAGTACCAACGCTGTGATCAATGTTTCATGATCATCTTCTGATCTCTCCCAATC 1564
QY 1309 AGAGGAGAGAGAGAGTCTGCTTCTCCGCGCAATGAATGAAGACAGCGGCTTCCCA 1368
Db 1565 AGAGGAGAGAGAGAGAGTCTGCTTCTCCGCGCAATGAATGAAGACAGCGGCTTCCCA 1624
QY 1369 TTTGTACAAAGTCAACCGCGTTTAAATCCAGAGGCTAGATGTGAGTGAAGCTTTCAG 1428
Db 1625 TTTGTACAAAGTCAACCGCGTTTAAATCCAGAGGCTAGATGTGAGTGAAGCTTTCAG 1684
QY 1429 CCGCGGGAAGATGAATTAATGATGCCCATTTAAGAAAGATGTCTGTGACAGCGGTGA 1488
Db 1685 CCGCGGGAAGTGA----- 1699
QY 1489 ATGGAGGTTTGGCGAGGACCGGCTCCAAAGATCTGGGTCAATGAGAGCAAGCAAGTGT 1548
Db 1700 -----GCAGAGCCTGAGCAATGCTATCTGGGTCAATGAGAGCAAGCAAGTGT 1747
QY 1549 GTACTTCCAGGGACCAAGAGACAGCGGCTGTGAGACACCACTCTAAGTGTGACTATGA 1608
Db 1748 GTACTTCCAGGGACCAAGAGACAGCGGCTGTGAGACACCACTCTAAGTGTGACTATGA 1807
QY 1609 GCGCGCGCGAGATGCTAGACCTCTACACAGCGCGGCTTCCCATGAGTGTCTCATAGAG 1668
Db 1808 GCGCGCGCGAGATGCTAGACCTCTACACAGCGCGGCTTCCCATGAGTGTCTCATAGAG 1867
QY 1669 CCAGAACTTGACATGTTGCTGACGCACTACAGAGCGGTGAGCAAGCGGCGCTTCCGTGA 1728
Db 1868 CCAGAACTTGACATGTTGCTGACGCACTACAGAGCGGTGAGCAAGCGGCGCTTCCGTGA 1927
QY 1729 CGTCTACAGCTGAGCGGCGCGGACGAGACCCCTGTGACAAAGAGCGCGCTTCTGGGC 1788
Db 1928 CGTCTACAGCTGAGCGGCGCGGACGAGACCCCTGTGACAAAGAGCGCGCTTCTGGGC 1987
QY 1789 TAGCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1848
Db 1988 TAGCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2000
QY 1849 CCACAGCGGCTCGAGTGTGGGCTCTACAGGATGATCTACAGGCCCAAGCGCTTTCAGCC 1908
Db 2021 CCACAGCGGCTCGAGTGTGGGCTCTACAGGATGATCTACAGGCCCAAGCGCTTTCAGCC 2080
QY 1909 AGGGAAGAGACCCACCGCTCTTGTATGAGAGGCGCCCAAGGTGACGCTGTGAA 1968
Db 2081 AGGGAAGAGACCCACCGCTCTTGTATGAGAGGCGCCCAAGGTGACGCTGTGAA 2140
QY 1969 TAACTCTTCAAGAGCATCAACTTGTGCGGCTCAACACACTGTGCTCCGCGCTACGC 2028
Db 2141 TAACTCTTCAAGAGCATCAACTTGTGCGGCTCAACACACTGTGCTCCGCGCTACGC 2200
QY 2029 CGTGTGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2088
Db 2201 CGTGTGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2260
QY 2089 GAAAGCAAAATGGGCAAGGTGAGATCGAGAGCAGAGTGAAGGCGCTGCAAGTGTGAGC 2148
Db 2261 GAAAGCAAAATGGGCAAGGTGAGATCGAGAGCAGAGTGAAGGCGCTGCAAGTGTGAGC 2320
QY 2149 CGAAGATATGAGCTTCAATCGACTGAGCGAGTGTGAGCATCATGATGATGATGATGATG 2208
Db 2321 CGAAGATATGAGCTTCAATCGACTGAGCGAGTGTGAGCATCATGATGATGATGATGATG 2380
QY 2209 CTTCTCTCGCTCATGGGGCTTAATCACAAGCCCAAGGTGTTCAAGGTGAGCATGCGGG 2268
Db 2381 CTTCTCTCGCTCATGGGGCTTAATCACAAGCCCAAGGTGTTCAAGGTGAGCATGCGGG 2440
|||||

```
OY 2269 TGCCTGACGCTGCTGATGAGCTGACGACAGGGGTACACTGAGCGCTACATGAGCGT 2328
    |||||||
DB 2441 TGCCTGACGCTGCTGATGAGCTGACGACAGGGGTACACTGAGCGCTACATGAGCGT 2500
OY 2329 CCCTGAGAACACACAGCAGCGCTATAGAGCGGGTTCCTGCGCTCAGCTGAGAAAGCT 2388
    |||||||
DB 2501 CCCTGAGAACACACAGCAGCGCTATAGAGCGGGTTCCTGCGCTCAGCTGAGAAAGCT 2560
OY 2389 GCCCAATGAGCCCAACCGCTTGTATCTCCACGCGCTTCTGAGAGAAAGCTGACCTT 2448
    |||||||
DB 2561 GCCCAATGAGCCCAACCGCTTGTATCTCCACGCGCTTCTGAGAGAAAGCTGACCTT 2620
OY 2449 TTTCCACACAAATTCCTGCTGCTCCCACTGATCCGAGAGAGGAAACCTTACAGCTC-- 2506
    |||||||
DB 2621 TTTCCACACAAATTCCTGCTGCTCCCACTGATCCGAGAGAGGAAACCTTACAGCTC 2680
OY 2507 -----CAGATCTACCCCAACGAGACACAGATATTCGCTG 2541
    |||||||
DB 2681 GGTGGCCCTGCTCTGCTGCTCCCGCAGATCTACCCCAACGAGAGACAGATATTCGCTG 2740
OY 2542 CCCCGAGTGGGGGAGCAGTATGAGTACGTTGCTGCACTTCTACAGGAATACCTG 2601
    |||||||
DB 2741 CCCCGAGTGGGGGAGCAGTATGAGTACGTTGCTGCACTTCTACAGGAATACCTG 2800
OY 2602 A 2602
    |
DB 2801 A 2801

RESULT 11
ABK83341
ID ABK83341 standard; cDNA; 4037 BP.
XX
AC ABK83341;
XX
DE 12-AUG-2002 (first entry)
XX
cDNA encoding human DPRP-2 splice variant #9.
XX
Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KM DPRP; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychiatric disorder; neurological disorder;
KM dyskinesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI; 2002-444178/47.
XX
P-PSDB; AB661610.
XX
New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure: Page 103-104; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
```

```
CC proteases referred to as dipeptidyl peptidase IV (DPP-IV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABK83342-ABK83343 encode human DPRP proteins.
SQ
Sequence 4037 BP; 869 A; 1268 C; 1131 G; 769 T; 0 other;
Query Match 84.4%; Score 2208.4; DB 24; Length 4037;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 1; Indels 182; Gaps 2;
OY 11 ATGGCCACCAACCGGAGCCCAACGCGCGAGCGAGCGACGCGCCACAGATGACCG 70
    |||||||
DB 367 ATGGCCACCAACCGGAGCCCAACGCGCGAGCGAGCGACGCGCCACAGATGACCG 426
OY 71 GCGCGCCGCTTCAGGTGAGAAAGCACTGCTGGAGCGGCTCCGAGCATCATCACGCG 130
    |||||||
DB 427 GCGCGCCGCTTCAGGTGAGAAAGCACTGCTGGAGCGGCTCCGAGCATCATCACGCG 486
OY 131 AGCCGCAATACCTCGGCGCTCATGTCACAAAGGCGCCCGACAGCTTCCAGTTGTGCGAG 190
    |||||||
DB 487 AGCCGCAATACCTCGGCGCTCATGTCACAAAGGCGCCCGACAGCTTCCAGTTGTGCGAG 546
OY 191 AAGACGATGAGTCTGGGCGCCCACTCCACGCGCTTACTACTGCGGAATGCGATGCG 250
    |||||||
DB 547 AAGACGATGAGTCTGGGCGCCCACTCCACGCGCTTACTACTGCGGAATGCGATGCG 606
OY 251 AGCCGAGAAACTCCCTCTCTACTCTGAGANTCCCAAGAGGTCGGAAGAAGGCTCTG 310
    |||||||
DB 607 AGCCGAGAAACTCCCTCTCTACTCTGAGANTCCCAAGAGGTCGGAAGAAGGCTCTG 666
OY 311 CCGCTCTGCTCCGGAAGAGAGTGTGATGATTCACAGGCGACCGCCACCATGGGGTC 370
    |||||||
DB 667 CCGCTCTGCTCCGGAAGAGAGTGTGATGATTCACAGGCGACCGCCACCATGGGGTC 726
OY 371 TACTCTCGGAGAGAGAGTGTGAGAGGAGCGAAAGCGCTGGGGTCTTTCGATCAC 430
    |||||||
DB 727 TACTCTCGGAGAGAGAGTGTGAGAGGAGCGAAAGCGCTGGGGTCTTTCGATCAC 786
OY 431 TCTTACGACTTCCACAGCAGAGAGTGGCTCTCTCTCTCCAGGCGACGAAACGCTCTTC 490
    |||||||
DB 787 TCTTACGACTTCCACAGCAGAGAGTGGCTCTCTCTCTCCAGGCGACGAAACGCTCTTC 846
OY 491 CACTGTGCGAGCGCGGCAAGACGGCTTCATGTGTCTCCCTATGAACCGCTGGAATTC 550
    |||||||
DB 847 CACTGTGCGAGCGCGGCAAGACGGCTTCATGTGTCTCCCTATGAACCGCTGGAATTC 906
OY 551 AAGACCCAGTGTCCAGGCGCCGAGTGAACCCCAATATGCGCTCCGACCTGCGCTTC 610
    |||||||
DB 907 AAGACCCAGTGTCCAGGCGCCGAGTGAACCCCAATATGCGCTCCGACCTGCGCTTC 966
OY 611 TTTCTCTTATCATTAACAGCAGCTGTGGTGGCCACATCGAGACGAGCGAGCGG 670
    |||||||
DB 967 TTTCTCTTATCATTAACAGCAGCTGTGGTGGCCCAACATCGAGACGAGCGAGCGG 1026
OY 671 CGGCTACCTTCTGCGACCAAGTTTATTCATGTCTCGATGACCCCAAGTCTGGGGGT 730
    |||||||
DB 1027 CGGCTACCTTCTGCGACCAAGTTTATTCATGTCTCGATGACCCCAAGTCTGGGGGT 1086
OY 731 GTGGCCACCTTCTGCTATAGAGAAAGATTGACCGCTTCACTGGATGACGTGTCGCC 790
    |||||||
DB 1087 GTGGCCACCTTCTGCTATAGAGAAAGATTGACCGCTTCACTGGATGACGTGTCGCC 1146
OY 791 ACAGCCTCTGTGGAAAGTTTCAGAGGGCTTCAGAGCCTGCGAATCTGTATGAGAAATC 850
```


Db 1147 ACAGCCTCCCTGGAGAGGTTGAGAGGGCTCAGAGCGCTGGAAATCCTGATGAGAGAGTC 1206
QY 851 GATGAGTCCAGAGGTGGAGGTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGAGAACAGC 910
Db 1207 GATGAGTCCAGAGGTGGAGGTCATTCACGTCCCTCTCTCTCGCTAGAGAAAGAGAACAGC 1266
QY 911 GACTCGATGAGTACCCAGAGAGAGAGAGAGATCCCAAGATTGCTTGAACCTGGCT 970
Db 1267 GACTCGATGAGTACCCAGAGAGAGAGAGAGATCCCAAGATTGCTTGAACCTGGCT 1326
QY 971 GAGTCCAGACTGACAGAGAGAGAGAGATGCTGACAGAGAGAGAGAGAGAGAGAGAGAG 1030
Db 1327 GAGTCCAGACTGACAGAGAGAGAGAGATGCTGACAGAGAGAGAGAGAGAGAGAGAGAG 1386
QY 1031 CCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCAGAGAGAGAGAGAGAGAGAGAGAG 1090
Db 1387 CCTTCAGCTCGCTGTTCCCGAAGGTGAGTACATCCAGAGAGAGAGAGAGAGAGAGAGAG 1446
QY 1091 GGCAGAAATACGCTGGGCGCATGTTCTTGACCGGCGCCAGAGAGTGGCTCCAGCTGCTCTC 1150
Db 1447 GGCAGAAATACGCTGGGCGCATGTTCTTGACCGGCGCCAGAGAGTGGCTCCAGCTGCTCTC 1506
QY 1151 CTCGCCCGGCGCTGTTCTCATCCGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
Db 1507 CTCGCCCGGCGCTGTTCTCATCCGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
QY 1211 AGAGCTGTCCCGAGAGATGTCACAGCCGTATGCTGACAGAGAGAGAGAGAGAGAGAGAG 1270
Db 1567 AGAGCTGTCCCGAGAGATGTCACAGCCGTATGCTGACAGAGAGAGAGAGAGAGAGAGAG 1626
QY 1271 ATCAATGTTATGATGATCATCTTCTATCCCTTCGCCCAATCAGAGAGAGAGAGAGAGAG 1330
Db 1627 ATCAATGTTATGATGATCATCTTCTATCCCTTCGCCCAATCAGAGAGAGAGAGAGAGAG 1686
QY 1331 TTTCTCCGCGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1390
Db 1687 TTTCTCCGCGCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1746
QY 1391 TTTAAATCCCAAGAGGCTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 1450
Db 1747 TTTAAATCCCAAGAGGCTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 1806
QY 1451 TGCCGCCATTAAGAGAGAGATGCTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1510
Db 1807 TGCCGCCATTAAGAGAGAGATGCTCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
QY 1511 GGCCTCAAGATCTGGGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1570
Db 1867 GGCCTCA-----AGGGCACCAGAGAG 1887
QY 1571 ACGCGCGTGGAG 1630
Db 1888 ACGCGCGTGGAG 1947
QY 1631 CTCACAGAGCGCGCTCTCCATAGCTGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1690
Db 1948 CTCACAGAGCGCGCTCTCCATAGCTGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAG 2007
QY 1691 AGCCACTACAG 1750
Db 2008 AGCCACTACAG 2067
QY 1751 GAG 1810
Db 2068 GAG 2127
QY 1811 TGCCGCCCGGATTAAGTCTCTCCAGAGATCTTCCATTTCCAGAGAGAGAGAGAGAGAGAG 1870
Db 2128 TGCCGCCCGGATTAAGTCTCTCCAGAGATCTTCCATTTCCAGAGAGAGAGAGAGAGAGAG 2187
QY 1871 CTCTAGGAGATGATCTCAAG 1930

Db 2188 CTCTAGGAGATGATCTCAAG 2247
QY 1931 CTCTAGGAGATGAG 1990
Db 2248 CTCTAGGAGATGAG 2307
QY 1991 TACTTGGCGGCTCAACACACTGAGGCTCCCTGGAGTACAGAGAGAGAGAGAGAGAGAGAGAG 2050
Db 2308 TACTTGGCGGCTCAACACACTGAGGCTCCCTGGAGTACAGAGAGAGAGAGAGAGAGAGAGAG 2367
QY 2051 GGCCTCTGTACAG 2110
Db 2368 GGCCTCTGTACAG 2427
QY 2111 GAGATGAG 2170
Db 2428 GAGATGAG 2487
QY 2171 CTGAGAGAGAGTGGCATTCATGAGGCTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2230
Db 2488 CTGAGAGAGAGTGGCATTCATGAGGCTGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2547
QY 2231 ATCCAG 2290
Db 2548 ATCCAG 2571
QY 2291 GGCCTAG 2350
Db 2572 ----- 2571
QY 2351 TATGAG 2410
Db 2572 -----GCCAAG 2584
QY 2411 CTATGCTTCAAG 2470
Db 2585 CTATGCTTCAAG 2644
QY 2471 TCCCAAG 2530
Db 2645 TCCCAAG 2704
QY 2531 AGTATGCTGCTCCCGAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2590
Db 2705 AGTATGCTGCTCCCGAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2764
QY 2591 GAATGCTGAGAG 2604
Db 2765 GAATGCTGAGAG 2778
RESULT 12
ABR83340
ID ABR83340 standard; cDNA; 4120 BP.
XX ABR83340;
AC
XX
DT 12-AUG-2002 (first entry)
XX
XX cDNA encoding human DPP-2 splice variant #8.
DE
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPP-2;
KW DPP-IV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder; gene; ss.
OS
XX Homo sapiens.
XX
XX
XX W0200231134-A2.

PD 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
XX
XX MPI: 2002-444178/47.
DR P-PSDB; ABG61609.
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
XX -
XX Disclosure: Page 100-101; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABR83322-ABR83343 encode human DPPR proteins.
XX
XX Sequence 4120 BP; 884 A; 1298 C; 1162 G; 776 T; 0 other:
XX
XX Query Match 84.4%; Score 2208.4; DB 24; Length 4120;
XX Best Local Similarity 92.9%; Pred. No. 0;
XX Matches 2411; Conservative 0; Mismatches 1; Indels 182; Gaps 2;
XX
OY 11 AAGGCGACACCGGGAGCCCAACGCGCGACGAGCGAGCGACGCCCGACATGACCGG 70
DB 367 ATGGCCACACCGGGAGCCCAACGCGCGACGAGCGAGCGACGCCCGACATGACCGG 426
OY 71 GCGCGCCGCTTCAGGTGCGAAGCACTGTGGAGCGGCTCGGAGCATCATCCAGCGC 130
DB 427 GCGCGCCGCTTCAGGTGCGAAGCACTGTGGAGCGGCTCGGAGCATCATCCAGCGC 486
OY 131 AGCCGCACTACTCGGGCTCATTTGTCAACAAGCGGCCCGACGACTTCCAGTTTGTGCG 190
DB 487 ACCCGCACTACTCGGGCTCATTTGTCAACAAGCGGCCCGACGACTTCCAGTTTGTGCG 546
OY 191 AAGAGGAGTACTGTGGGCGCCACTCCGCGCTACTACTGGGAGTGCATTTGGC 250
DB 547 AAGAGGAGTACTGTGGGCGCCACTCCGCGCTACTACTGGGAGTGCATTTGGC 606
OY 251 AGCCGAGAACTCCCTCCTTACTCTGAGATTCCCAAGAGTCCGGAAGAAGGCTGTG 310
DB 607 AGCCGAGAACTCCCTCCTTACTCTGAGATTCCCAAGAGTCCGGAAGAAGGCTGTG 666
OY 311 CTGCTCTCTCTCTGGAAGAGATGCTGATTCATTTCCAGGCGACGCCCGACCATGGGGTC 370
DB 667 CTGCTCTCTCTCTGGAAGAGATGCTGATTCATTTCCAGGCGACGCCCGACCATGGGGTC 726
OY 371 TACTCTCGGAGAGAGAGTGTGAGGAGCGGAAAGCGCTGGGGTCTTCCGATACAC 430
DB 727 TACTCTCGGAGAGAGTGTGAGGAGCGGAAAGCGCTGGGGTCTTCCGATACAC 786
OY 431 TCTTACGACTTCCACAGCAGAGTGGCTTCTCTCTTCCAGGCGACGAAAGCGCTTTC 490
DB 787 TCTTACGACTTCCACAGCAGAGTGGCTTCTCTCTTCCAGGCGACGAAAGCGCTTTC 846

OY 491 CACTGTGCGAGCGGCGGCAAGAAAGGCTTTCATGTGTGCTTCCCTATGAAACCGCTGAAATC 550
DB 847 CACTGTGCGAGCGGCGGCAAGAAAGGCTTTCATGTGTGCTTCCCTATGAAACCGCTGAAATC 906
OY 551 AAGACCACTGCTCAGAGGCGCCGGATGAGACCCCAAAATCTGCCCTCGCACTGCTTTC 610
DB 907 AAGACCACTGCTCAGAGGCGCCGGATGAGACCCCAAAATCTGCCCTCGCACTGCTTTC 966
OY 611 TTCTCTTCATCAATTAACAGCGACTGTGGGTGGCCCAATGAGAGAGCGGAGAGCGG 670
DB 967 TTCTCTTCATCAATTAACAGCGACTGTGGGTGGCCCAATGAGAGAGCGGAGAGCGG 1026
OY 671 CGGCTACCTTCTGCGACCAAGGTTTATCAATGTCTGTGATGACCCCAAGTCTGGGGT 730
DB 1027 CGGCTACCTTCTGCGACCAAGGTTTATCAATGTCTGTGATGACCCCAAGTCTGGGGT 1086
OY 731 GTGGCCACTTGTCTATACAGGAAGATTGACCGCTTCACTGGGTACTGTGTGCTCC 790
DB 1087 GTGGCCACTTGTCTATACAGGAAGATTGACCGCTTCACTGGGTACTGTGTGCTCC 1146
OY 791 ACAGCTCTCGGAAAGTTTCAGAGGCTCAGAGGCTGCGAATCCTGTATGAGAAATC 850
DB 1147 ACAGCTCTCGGAAAGTTTCAGAGGCTCAGAGGCTGCGAATCCTGTATGAGAAATC 1206
OY 851 GATGACTCCGAGGTGAGAGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAAAGAGAGC 910
DB 1207 GATGACTCCGAGGTGAGAGTCAATTCAGTCCCTCTCTGCGCTAGAAAGAAAGAGAGC 1266
OY 911 GACTGTATTCGTTACCCAGAGACAGCAGCAAGAAATCCCAAGTTTCTTGAATCTGCT 970
DB 1267 GACTGTATTCGTTACCCAGAGACAGCAGCAAGAAATCCCAAGTTTCTTGAATCTGCT 1326
OY 971 GAGTTCCAGACTGACAGCGACGAGCAAGATGCTGTGACCCAGAGAAAGAGAGTGGTGCAG 1030
DB 1327 GAGTTCCAGACTGACAGCGACGAGCAAGATGCTGTGACCCAGAGAAAGAGAGTGGTGCAG 1386
OY 1031 CCGTTACGCTCGCTGTCCCGAAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1090
DB 1387 CCGTTACGCTCGCTGTCCCGAAGGTGAGTACATGCCAGGCGCGGTGAGACCGCGGAT 1446
OY 1091 GCGAATACGCTCGGCGCAATTTCTGTGACCGCGCCAGCAATGCTTCCAGTCTGCTCTC 1150
DB 1447 GCGAATACGCTCGGCGCAATTTCTGTGACCGCGCCAGCAATGCTTCCAGTCTGCTCTC 1506
OY 1151 CTCCCGCCGCGCTGTTCATCCGAGACAGAGAAATGAGAGACAGCGGCTTACCTTGC 1210
DB 1507 CTCCCGCCGCGCTGTTCATCCGAGACAGAGAAATGAGAGACAGCGGCTTACCTTGC 1566
OY 1211 AAGCTGTCCCAAGAAATTCAGCCGTAATGTGTGTACAGAGAGTCAACCAAGTCTG 1270
DB 1567 AAGCTGTCCCAAGAAATTCAGCCGTAATGTGTGTGTACAGAGAGTCAACCAAGTCTG 1626
OY 1271 ATCAATGTTCATGACATCTTATCCTTCCCAATCGAAGAGAGAGAGAGAGTCTGC 1330
DB 1627 ATCAATGTTCATGACATCTTATCCTTCCCAATCGAAGAGAGAGAGAGAGTCTGC 1686
OY 1331 TTTTCTCCGCGCAATGAAATGACAGCGGCTCTGCGATTTGTACAAATTCACCGCGTT 1390
DB 1687 TTTTCTCCGCGCAATGAAATGACAGCGGCTCTGCGATTTGTACAAATTCACCGCGTT 1746
OY 1391 TTTTAAATTCAGAGGCTCAGATTGAGAGTCCCTTACGCGCGGAGAGATATTTAAG 1450
DB 1747 TTTTAAATTCAGAGGCTCAGATTGAGAGTCCCTTACGCGCGGAGAGATATTTAAG 1806
OY 1451 TGCCTCATTAAGAGAGATTTGCTCTGACACAGCGGTGAATGGAGAGTTTGGCAGGAC 1510
DB 1807 TGCCTCATTAAGAGAGATTTGCTCTGACACAGCGGTGAATGGAGAGTTTGGCAGGAC 1866
OY 1511 GCGTCAAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGACCAAGAGC 1570
DB 1867 GCGTCA-----AGGCGACCAAGAGC 1887
OY 1571 AGCGCGTGGAGACCACTACTAGTGTGTCAGTATGAGCGCGCGGAGATCTGATCC 1630

Db	1888	ACGGCGGTGGAGACCAACCTCTACGTGSHACACTATGAGGGGCGGGAGATCGTAACG	1947
Qy	1631	CTACACCAACGCGGGCTTCTCCATAGCTGCTCATGAGCCAGACCTTGACATGTTGCTC	1650
Db	1948	CTCACCAACGCGCGGCTTCTCCATAGCTGCTCATGAGCCAGACCTTGACATGTTGCTC	2007
Qy	1691	AGGCACATACAGCAAGCGTGAAGCAAGCGCGGCTCGGCGTGAACGTCTACAACTGAGGGGCCCC	1750
Db	2008	AGGCACATACAGCAAGCGTGAAGCAAGCGCGGCTCGGCGTGAACGTCTACAACTGAGGGGCCCC	2067
Qy	1751	GAGCAGACACCCCTGACACAAGCAGCCCGCTCTTGGGCTAGCATGATGAGGAGCCAGC	1810
Db	2068	GAGGAGACACCCCTGACACAAGCAGCCCGCTCTTGGGCTAGCATGATGAGGAGCCAGC	2127
Qy	1811	TGCCCCCGGATTAATGTTCTCTCCAGAGATCTTCATTTCCACAGCGCTCGATGTGCGG	1870
Db	2128	TGCCCCCGGATTAATGTTCTCTCCAGAGATCTTCATTTCCACAGCGCTCGATGTGCGG	2187
Qy	1871	CTCTAGGGCAATGATTCACAAGCCCCCAAGGCTTCGACGCCAGGGAAGAACCCCAACGCTC	1930
Db	2188	CTCTAGGGCAATGATTCACAAGCCCCCAAGGCTTCGACGCCAGGGAAGAACCCCAACGCTC	2247
Qy	1931	CTCTTTGTATATGAGAGCGCCCCAGGTGACAGCTGAGTAATTAATCTCTTAAAGAGCATCAAG	1990
Db	2248	CTCTTTGTATATGAGAGCGCCCCAGGTGACAGCTGAGTAATTAATCTCTTAAAGAGCATCAAG	2307
Qy	1991	TACTTGGCGCTCAACACACTGAGGCTCTCCGTGGCTACAGCCGCTGGTTGTATTTGACGGCAGG	2050
Db	2308	TACTTGGCGCTCAACACACTGAGGCTCTCCGTGGCTACAGCCGCTGGTTGTATTTGACGGCAGG	2367
Qy	2051	GGCTCCTCTGACGAGAGGGGCTTCGGTTTCGAAGGGGGCCCTAAAAACCAATGAGCCAGGTG	2110
Db	2368	GGCTCCTCTGACGAGAGGGGCTTCGGTTTCGAAGGGGGCCCTAAAAACCAATGAGCCAGGTG	2427
Qy	2111	GAGATCGAGGACCAAGGTTGAGGGGCTTCGAGTTTCGTGGCCGAGAAAGTATGGCTCATGTACAC	2170
Db	2428	GAGATCGAGGACCAAGGTTGAGGGGCTTCGAGTTTCGTGGCCGAGAAAGTATGGCTCATGTACAC	2487
Qy	2171	CTGAGCCGAGTTGCCATCATATGCGTGTCTCTAAGGGGGCTTCCTGCTCATGAGGGCTA	2230
Db	2488	CTGAGCCGAGTTGCCATCATATGCGTGTCTCTAAGGGGGCTTCCTGCTCATGAGGGGGCTA	2547
Qy	2231	ATCCACAAGCCCCAGGTGTTCAAGGTGCGCATTCGCGGGTTCGCCGGTCAACGCTCTGATG	2290
Db	2548	ATCCACAAGCCCCAGGTGTTCAAG-----	2571
Qy	2291	GCGTACGACACAGGGTCACTAGAGGCTCATATGAGAGTCCCTGAGAACACACAGCAGCGC	2350
Db	2572	-----	2571
Qy	2351	TATGAGCGGGGTTCCGCGGCTCTGCACGTGAGAGAACTGCCAATGAGGCCCAACCGCTTG	2410
Db	2572	-----GCCACACGCTTG	2584
Qy	2411	CTTATTCCTCACAGGCTTCTGAGAGAAAAGTGCACATTTTTCCACACAATCTCTCGTC	2470
Db	2585	CTTATTCCTCACAGGCTTCTGAGAGAAAAGTGCACATTTTTCCACACAATCTCTCGTC	2644
Qy	2471	TCCCAACTGATCCGAGAGAGGAAACCTTACAGACTCCAGATCTTACCCTCAAGAGAGAC	2530
Db	2645	TCCCAACTGATCCGAGAGAGGAAACCTTACAGACTCCAGATCTTACCCTCAAGAGAGAC	2704
Qy	2531	AGTATTTGCGTCCCGAGTGGGGCGAGCACTATGAAGTACAGTTCGTGACTTTTACAG	2590
Db	2705	AGTATTTGCGTCCCGAGTGGGGCGAGCACTATGAAGTACAGTTCGTGACTTTTACAG	2764
Qy	2591	GAATACCTCTGAGC	2604
Db	2765	GAATACCTCTGAGC	2778

ID	AA157880	AA157880 standard; cDNA: 3262 BP.
XX	AA157880;	
XX	22-OCT-2001 (first entry)	
XX	Human polynucleotide SEQ ID NO 83.	
XX	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
XX	amortrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
XX	chemokine; thrombolytic; drug screening; arthritis; inflammation;	
XX	leukaemia; ss.	
XX	Homo sapiens.	
XX	MO200153312-A1.	
XX	26-JUL-2001.	
XX	26-DEC-2000; 2000WO-US34263.	
XX	21-JAN-2000; 2000US-0488725.	
XX	25-APR-2000; 2000US-0552317.	
XX	09-JUL-2000; 2000US-0598042.	
XX	19-JUL-2000; 2000US-0620312.	
XX	03-AUG-2000; 2000US-0653450.	
XX	14-SEP-2000; 2000US-0662191.	
XX	19-OCT-2000; 2000US-0693036.	
XX	29-NOV-2000; 2000US-0727344.	
XX	(HYSE-) HYSEQ INC.	
XX	Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
XX	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX	WPI: 2001-442253/47.	
XX	P-PSDB; AAM38724.	
XX	Novel nucleic acids and polypeptides, useful for treating disorders	
XX	such as central nervous system injuries -	
XX	Claim 1; SEQ ID NO 83; 10078pp; English.	
XX	The invention relates to human nucleic acids (AA157798-AA161369) and	
XX	the encoded polypeptides (AAM38642-AAM42213) with neotropic,	
XX	immunosuppressant and cytostatic activity. The polynucleotides are useful	
XX	in gene therapy. A composition containing a polypeptide or polynucleotide	
XX	of the invention may be used to treat diseases of the peripheral nervous	
XX	system, such as peripheral nervous injuries, peripheral neuropathy and	
XX	localised neuropathies and central nervous system diseases, such as	
XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
XX	utilisation of the activities such as: Immune system suppression,	
XX	Activin/inhibin activity, chemotactic/chemokine activity, haemostatic	
XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
XX	assays for receptor activity, arthritis and inflammation, leukaemia and	
XX	C.N.S disorders.	
XX	Note: The sequence data for this patent did not form part of the printed	
XX	specification.	
XX	Sequence 3262 BP: 687 A; 1019 C; 931 G; 625 T; 0 other;	
XX	Query Match 79.9%; Score 2091.4; DB 22; Length 3262;	
XX	Best Local Similarity 94.4%; Pred. No. 0;	
XX	Matches 2256; Conservative 0; Mismatches 21; Indels 112; Gaps 4;	
XX	243 CATATGCGAGCGGAGAGAACTCCCTCTTACTCTGAGATTTCCCAAGAAGTCCGGAAG 302	
XX	33 CATATGCGAGCGGAGAGAACTCCCTCTTACTCTGAGATTTCCCAAGAAGTCCGGAAG 91	

QY	303	AGGCTCTGCTGCTCTCTCTCTCGAAGACAGATGCTGATCATTTTCCAGGCTACGCCAC	362
Db	92	AGGCTCTGCTGCTCTCTCTCTCTCGAAGACAGATGCTGATCATTTTCCAGGCTACGCCAC	151
QY	363	ATGGGGTCTACTCTCGGGAGAGAGAGCTGTGAGGGAGAGGAAAGCCCTGGGGGTCTTG	422
Db	152	ATGGGGTCTACTCTCGGGAGAGAGAGCTGTGAGGGAGAGGAAAGCCCTGGGGGTCTTG	211
QY	423	GCATCACCTCTCTACGACTTCCACAGCGAGAGTGGCTCTTCTCTTCAGGCCACACAA	482
Db	212	GCATCACCTCTCTACGACTTCCACAGCGAGAGTGGCTCTTCTCTTCAGGCCACACAA	271
QY	483	GCCCTTTCACACTGTGTCGGAGGGGGGCAAGAAAGGCTTCATGGGTGTCCTATGAAACCG	542
Db	272	GCCCTTTCACACTGTGTCGGAGGGGGGCAAGAAAGGCTTCATGGGTGTCCTATGAAACCG	331
QY	543	TGGAATTAAGAACCCAGTGTCTCAAGGGCCCCGATGAGACCCAAATATCTCCCTGCCGAC	602
Db	332	TGGAATTAAGAACCCAGTGTCTCAAGGGCCCCGATGAGACCCAAATATCTCCCTGCCGAC	391
QY	603	CTGCGTCTTCTCTCTTATATTAAGACGACCTGTGGGTGGCCAAATCTGAGACAGGCG	662
Db	392	CTGCGTCTTCTCTCTTATATTAAGACGACCTGTGGGTGGCCAAATCTGAGACAGGCG	451
QY	663	AGGAGCGGCGGCTGACCTTCTGGCACCAAGATTATCCATGTCTCGATGACCCCAAGT	722
Db	452	AGGAGCGGCGGCTGACCTTCTGGCACCAAGATTATCCATGTCTCGATGACCCCAAGT	511
QY	723	CTGCGGGTGTGGCCACTTGTGTATACAGAGAGAGTTTACACCGCTTCACTGGGTACTG	782
Db	512	CTGCGGGTGTGGCCACTTGTGTATACAGAGAGAGTTTACACCGCTTCACTGGGTACTG	571
QY	783	GGTCCGCCACAGCCCTCTGGGAAAGTTTCAAGGGCCCTCAAGAGCGTGCATTCGTATG	842
Db	572	GGTCCGCCACAGCCCTCTGGGAAAGTTTCAAGGGCCCTCAAGAGCGTGCATTCGTATG	631
QY	843	AGGAAGTCGATGATGATCCGAGGTGGAAGTCAATTCAGCTCCCTCTCTCGCTAGGAAGAA	902
Db	632	AGGAAGTCGATGATGATCCGAGGTGGAAGTCAATTCAGCTCCCTCTCTCGCTAGGAAGAA	691
QY	903	GGAGACGAGACTCGTATTCGGTATCCCCACAGACAGGACAGAAATCCCAAGATTGCCTTGA	962
Db	692	GGAGACGAGACTCGTATTCGGTATCCCCACAGACAGGACAGAAATCCCAAGATTGCCTTGA	751
QY	963	AACGTGCTGATGTTCCAGACTGACAGCGACGAGGCAAGATTCGTCGACCAGAGAGAGAC	1022
Db	752	AACGTGCTGATGTTCCAGACTGACAGCGACGAGGCAAGATTCGTCGACCAGAGAGAGAC	811
QY	1023	TGTGTACAGCCCTTACGTCGCTGTTTCCCAAGGTGGAATACATCGCCAGAGCCGGGTGGA	1082
Db	812	TGTGTACAGCCCTTACGTCGCTGTTTCCCAAGGTGGAATACATCGCCAGAGCCGGGTGGA	871
QY	1083	CCCGGATGAGCAATATACGCTGGGACCATGTTCTGTAGCCGGCCCCAGAGAGTGGCTCAGC	1142
Db	872	CCCGGATGAGCAATATACGCTGGGACCATGTTCTGTAGCCGGCCCCAGAGAGTGGCTCAGC	931
QY	1143	TGTGTCTCTCCCGCGGCGCTTTCATCCGACGACAGAGAGATGAGAGCGGGTAG	1202
Db	932	TGTGTCTCTCCCGCGGCGCTTTCATCCGACGACAGAGATGAGAGCGGGTAG	991
QY	1203	CCCTGTGCACAGCGTCCCGCAGGAATGTCCAGCGGATGTGGTATACAGAGAGGTCCACA	1263
Db	992	CCCTGTGCACAGCGTCCCGCAGGAATGTCCAGCGGATGTGGTATACAGAGAGGTCCACA	1051
QY	1263	ACGCTGTGATCAATGTTCATGACATCTTATCCCTTCCCCCAATCAGAGGAGAGAGAGC	1322
Db	1052	ACGCTGTGATCAATGTTCATGACATCTTATCCCTTCCCCCAATCAGAGGAGAGAGAGC	1111
QY	1323	AGCTCTGCTTTCTCCGCGCAATGAATGCAAGACGGCTTCTGCCATTTGTACAAATGCA	1382
Db	1112	AGCTCTGCTTTCTCCGCGCAATGAATGCAAGACGGCTTCTGCCATTTGTACAAATGCA	1171

OY	1383	CCGCCGTTTAAATFCCAGGCGCTACGATTGGAGTGAAGCCCTTCAGGCCCGGGGGAAGATG	144
Db	1172	CCGCCGTTTAAATFCCAGGCGCTACGATTGGAGTGAAGCCCTTCAGGCCCGGGGGAAGATG	1231
OY	1443	AATTAAAGTGGCCCATTAAGGAAGATTGCTGTGACCAGCGGTGAAATGGAGGTTTGG	1502
Db	1232	A-----GC	1233
OY	1503	CGAGGACAGGCGCTCCAMAGATCTGGGTCAAATGAGGAGACCAAGCTGGTGTACTTTCAGAGGCA	1563
Db	1235	AGAGCGCTGACAGTAATGATCTGGGTCAAATGAGGAGACCAAGCTGGTGTACTTTCAGAGGCA	1294
OY	1553	CCAAAGGACAGCGCGCTGGAGCACCACTCTACGTGGTCAGCTATATGAGGGCGCGCGCAGA	1622
Db	1295	CCAAAGGACAGCGCGCTGGAGCACCACTCTACGTGGTCAGCTATATGAGGGCGCGCGCAGA	1354
OY	1623	TCGTACGCGCTACACCAAGCCCGCGCTTCCCATATGCTGCATATGAGACCAGAACTTCGACA	1688
Db	1355	TCGTACGCGCTACACCAAGCCCGCGCTTCCCATATGCTGCATATGAGACCAGAACTTCGACA	1414
OY	1683	TGTTTCGTACGCACTACAGAGAGCGGTGAGACGCGCGCCCTGCGTCACGCTTACAAAGCTGA	1742
Db	1415	TGTTTCGTACGCACTACAGAGAGCGGTGAGACGCGCGCCCTGCGTCACGCTTACAAAGCTGA	1474
OY	1743	GGCGCCCCGAGAGAGACCCCTCGACAAAGAGCCCGCTTCTGGGCTAGCATGATGAGAG	1802
Db	1475	GGCGCCCCGAGAGAGACCCCTCGACAAAGAGCCCGCTTCTGGGCTAGCATGATGAGAGAG	1534
OY	1803	CAGCGAGCTGGCCCCCGGATTATGTTCTCAAGATATCTTCATTTTCCACAGCGGCTGG	1862
Db	1535	CAGCCA-----AGATCTTCCATTTCCACAGCGGCTGG	1567
OY	1863	ATGTGCGGCTCTACGGCATGATCTACAAAGCCCGACGCTTGCAGCGCAGGGAAGAAAGCAC	1922
Db	1568	ATGTGCGGCTCTACGGCATGATCTACAAAGCCCGACGCTTGCAGCGCAGGGAAGAAAGCAC	1622
OY	1923	CCACCGCTCTCTTGTATATGAGAGCCCGCAGGTGACGTGTAATTAATCTCTTCAAG	1982
Db	1628	CCACCGCTCTCTTGTATATGAGAGCCCGCAGGTGACGTGTAATTAATCTCTTCAAG	1687
OY	1983	GCATCAAGTACTTCCGGCTCAACAACATGGCCCTCGGGGTCAGCGCGGTGGTGTGATTG	2042
Db	1688	GCATCAAGTACTTCCGGCTCAACAACATGGCCCTCGGGGTCAGCGCGGTGGTGTGATTG	1747
OY	2043	ACGGCAGGGGCTCTCTGCAGCGAGGCTTCGTTCCAAAGGGGCCCTGAAAAACCAAAATGG	2102
Db	1748	ACGGCAGGGGCTCTCTGCAGCGAGGCTTCGTTCCAAAGGGGCCCTGAAAAACCAAAATGG	1807
OY	2103	GCCAGGTGAGAGATGAGAGACCAAGGTGAGAGGCTGCAATTCGTGGCCGGAAGATATGCGCT	2162
Db	1808	GCCAGGTGAGAGATGAGAGACCAAGGTGAGAGGCTGCAATTCGTGGCCGGAAGATATGCGCT	1867
OY	2163	TCATCGACCTGAGAGCGAGTGGCATCATGAGCGTGCATCAGAGGGGGCTTCCGTCTGGCGCA	2222
Db	1868	TCATCGACCTGAGAGCGAGTGGCATCATGAGCGTGCATCAGAGGGGGCTTCCGTCTGGCGCA	1927
OY	2223	TGGGGCTAATTCACAACCCCCAGGTTTCAAGGTGGCCATTCGCGGGTCCCGCGTCAACG	2282
Db	1928	TGGGGCTAATTCACAACCCCCAGGTTTCAAGGTGGCCATTCGCGGGTCCCGCGTCAACG	1987
OY	2283	TCGTGGATGGCCTACGACACAGGCTTACACTGAGCGCTACATGAGACCTCCCTGGAACAAC	2342
Db	1988	TCGTGGATGGCCTACGACACAGGCTTACACTGAGCGCTACATGAGACCTCCCTGGAACAAC	2047
OY	2343	AGCAGCGCTATGAGCGCGGTTCCGTGGCCCTGACCTGGAAGAGTGCCCAATGTGACCA	2402
Db	2048	AGCAGCGCTATGAGCGCGGTTCCGTGGCCCTGACCTGGAAGAGTGCCCAATGTGACCA	2107
OY	2403	ACCGCTTGCTTAATCTCCACAGGCTTCTGAGACAAAACGTGACTTTTTCACACAAACT	2462
Db	2108	ACCGCTTGCTTAATCTCCACAGGCTTCTGAGACAAAACGTGACTTTTTCACACAAACT	2167
OY	2463	TCCTCGCTCCCAACTGATCCGACAGGGAACCTTACAGCTC-----	2506

QY	1099	GCCTGGGCAATGTTTCTGTGGACGCGGCCCCACAGCATGTGGCTTCACGTGTCTCTCTCCCCC	1158
Db	1107	TTGGCTGGGCCCAATGTTTCTGTGGACGCTCCCAAGCAACGGCTTTCAGCTGTCTCTCTCCCCC	1168
QY	1159	GGCCCTGTTTCAATCCCGAGACAGGAATGTAGGAGCAGCGGCAATAGCCTCTGCCAGAGCT	1218
Db	1167	TGCTCTCTTCAATCCCGGCGCTGTGAGGTAGGCGCCAGCGGCAAGCAGCTGCCAGAGCGCT	1228
QY	1219	CCCCAGAAATGTCCAGCCGCTAATGTGGTGTGTCAGAGAGGTCCACCAACGTTGTGGATCAATGT	1278
Db	1227	CCCCAAGAAATGTGCAAGCCCTTTGTGCATCTATGAAGAAGTCCACCAATGTGTGGATCAACGT	1286
QY	1279	TCATGTACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGACAGACTCTGCTTCTCG	1338
Db	1287	CCAGACACATCTTCCACCCTGTTTCTCCAGGCTGAGGCGCCAGCAGGACTTGTGTTTCTTCG	1346
QY	1339	CGCCAAATGAATGCAGAGACCGGGCTTGGCCCTTGTGTCAAAAGTACCGCGCTTTTAAATTC	1398
Db	1347	TGCCAAGCAATGCAGAGACTGGCTTCTGCACTGTACAGAGGTACACGTGGAACCTTAAATC	1406
QY	1399	CCAGGGCTACGATTTGAGTGAAGCCCTTTCAGCCCGGGGGAAGATGTAATTAATGGCCCAT	1458
Db	1407	CAAGGACTATGACGTGAGACGGAAACCCCTACACCCCTACAGAAAGGTGAATTAATGGCCCAT	1466
QY	1459	TAAAGAGAGATTTCTCTGTACACAGCGGTGANTGGAGGTTTTGGCGAGGACGCGCTCAA	1518
Db	1467	CAAGAGAGAGGTCTCCCTGTGACAGTGGCGAGTGGGAGGCTTGTGGAGGACATGCGTCCAA	1526
QY	1519	GATCGGGGTCAATGAGGAGACCAAGCGGTGTCTTCCAGGGCCACCAAGAGACGCGCCT	1578
Db	1527	GATCTGGGTCAACGAGAGACGAAGAGCTGGTGTACTTTCAGGTACAAAGACACACCGCT	1586
QY	1579	GGAGCACCACTCTACCTGTGGTGTAGACTATGAGGCGGCGCGAGATCGTACGCTACAC	1638
Db	1587	GGAAATCACTCTATGTGTGTACAGTACAGAGTACAGAGCGAGATGTGTGGCTACAC	1646
QY	1639	GCCCGGCTTCTCCCATAGTGTCTTCATGAGCCAGAACTTGACATGTTGTGACGCACTA	1698
Db	1647	GCTGGGCTTCTCCACAGCTGTCTCATAGACCAGACCTTCGACATGTTGTGTAGCACTA	1706
QY	1699	CAGAGGCTGAGCAGCGCGCGCGCGAGCTCAAGAGCTGAGCGGCGCCGAGACGA	1758
Db	1707	CAGAGGTGTGACAGCGCACCCCTGTGTACATGTGTCAAGCTGTAGCGGCGCCGATGATGA	1766
QY	1759	CCCCCTGCACAGACGACCCCGCTTCTGGGTGTACAGATGAGGAGCAGCAGCTGCCCCC	1818
Db	1767	CCCACTGCACAGACACAGCGCTTCTGGGCGACCATGATGAGGACGACCAATTTGCCCCC	1826
QY	1819	GGATTAATGTTCTCTCAGAGATCTTCCATTTCCACACGCGCTCGAGTGTGGGCTCTACGG	1878
Db	1827	AGACTATGTGTGCCCCCTAGATCTTCCACTTTCACACCCCTGCAGAGCTGTGAGCTCTACGG	1886
QY	1879	CATCATGTACAGAGCCCCAGCGCTTGTGAGCCAGGGAAGAGACCCCAACGCTCTTTGT	1938
Db	1887	CATGATCTACAGAGCCACACACCCCTGCACTTGGAGGAAGACCCCACTGTCTTTGT	1946
QY	1939	ATATGAGAGGCCCCAGGTGTGAGCTGTGTGATATACCTCTTCAAGGACTCAAGTACTTGG	1998
Db	1947	CTATGTGGGGCCACAGAGGTGAGTTGTGGAACAACACTCTTTAAGGCACTCAATATCTGCG	2006
QY	1999	GCTCAACACACTGTGCTCTCCCTGGGCTACGCGGTGTGTGATTTGACGGCAGGGGCTCTG	2058
Db	2007	GCTAAATATCACTGTGCATCTTGTGGGCTATGCTGTGGTGTGATGTGATGTCGGGGCTCTG	2066
QY	2059	TCACGAGAGGCTTCGTGTGCAAGGGGCCCTGAAAAACCAATATGGGCAAGGTGTGATTCGA	2118
Db	2067	TCACGCGGGGCTGTGACTTTCAGAGGGGGCCCTGAAAAATCAAAATGGGCCAGGTGTGATTCGA	2126
QY	2119	GGACACAGGTGAGGGGCTGTGATTTGCGGCGCGAAGATATGGCTTCATTCGACCTGTAGCG	2178
Db	2127	GGACACAGGTGAGAGGCTTGTGCAAGTACGTGGCTGTGAGAGTATGTGCTTCATTTGACTGTAGCG	2186
QY	2179	AGTTGCATCATGCTGCTGTGCTACAGGGGGCTTCTCTCTGCTCATAGGGGCTCAATATCCAA	2238

Db	2187	 AGTGCACATTCATGGCTGAGTCTACAGGCGGCTTCCCTACTCATCATGGGCTCATCCACAA	2246
Qy	2239	 GCCCGAGGTGTTCAAGGTGGCCATTCGGGGGTGCCCGGTACACGTCGTGGATGGCCTACGA	2298
Db	2247	 GCCACAAATGTGTTCAAGGTATCCATTCGGGGGCGTCTCTGACATGTGTGGATGGCCTATGA	2306
Qy	2299	 CACAGGTACACTAAGGCTACATATGGACGTCCTTGAGAACACAGCAGCGCTATGAGGC	2358
Db	2307	 CACAGGTACACAGGAACGATACATGGATGTGTCGCCGAAATTAACACAGCAGCTATGAGGC	2366
Qy	2359	 GGGTTCCGTGGCCCTGACGCTGGAGAAAGCTGCCAATGAAGCCCAACGCTTGCCTTATCCT	2418
Db	2367	 AGGGTCTGTAGCCCTGCATGTGGAGAAAGCTGCCCAATGAGCTTAACCGCGCTGCTTATCCT	2426
Qy	2419	 CCAGGCGTCTCTGACGCAAAAGCTGCACCTTTTCCACACAAACTTCCTGCTGCCAACT	2478
Db	2427	 CCAGCGGCTTCCCTGGACGACGAAGACCTTCACTTCTTCCACAAATTTCTCGGTGTCACACT	2486
Qy	2479	 GATCCGAGCAGAGGAACCTTACAGGCTCCAGATCTACCCCAACGAGAGACACAGATTTCCG	2538
Db	2487	 GATCCGAGCAGGAAGCCATACCAAGCTTCAGATCTACCCCAACGAGAGACATAGATTCG	2546
Qy	2539	 CTGCGCCGAGTCCGGGCGAGCACTATGAAGTACGTTGCTGCACCTTCTACAGGAATACT	2598
Db	2547	 CTGCGCGGAGTCCGGAGAGGATTAACGAGGTGACGCGTGTGCACCTTCTTGCAGGAACACT	2606
Qy	2599	CTGAGC 2604 	
Db	2607	GTGAGC 2612	

RESULT	15
AAC75835	
ID	AAC75835 standard; cDNA: 2461 BP.
XX	
AC	AAC75835;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF1390 polynucleotide sequence SEQ ID NO:2779.
XX	
KM	Human; open reading frame; ORFX: detection; cytostatic; hepatotropic;
KM	humeral; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KM	anticonvulsant; osteophilic; antiarthritic; immunosuppressant; cardiac;
KM	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KM	antifungal; antihypertensive; antihemetic; antithyroid;
KM	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM	cholesterol ester storage; systemic lupus erythematosus; infection;
KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	MO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PE	31-MAR-2000; 2000MO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;


```

QY 1610 GCGGCCGCGGATGCTGACGCTCACCACGCCGCTTCTCCATAGCTGCTCCATGAGC 1669
    |||||
Db 2134 GCGGCCGCGGAGATCGTACGCTCACCACGCCGCTTCTCCATAGCTGCTCCATGAGC 2193
    |||||
QY 1670 CAGAACTTCGACATGTTGTCAGCCACTACAGCAGCGTGAAGCAGCGCCCTGGGTGAC 1729
    |||||
Db 2194 CAGAACTTCGACATGTTGTCAGCCACTACAGCAGCGTGAAGCAGCGCCCTGGGTGAC 2253
    |||||
QY 1730 GTCACAAAGCTGAGCGGCGCCGACGAGACCCCTGCACAAGCAGCCCGCTTCTGGGCT 1789
    |||||
Db 2254 GTCACAAAGCTGAGCGGCGCCGACGAGACCCCTGCACAAGCAGCCCGCTTCTGGGCT 2313
    |||||
QY 1790 AGCATGATGAGAGCAGCAGCTGCCCCCGGATTATGTTCTCCAGAGATCTTCATTTC 1849
    |||||
Db 2314 AGCATGATGAGAGCAGCAGCTGCCCCCGGATTATGTTCTCCAGAGATCTTCATTTC 2373
    |||||
QY 1850 CACACGGGCTCGGATGTCGGGCTCTACGGCATGATCAAGCCCGCAGGCTTGACAGCA 1909
    |||||
Db 2374 CACACGGGCTCGGATGTCGGGCTCTACGGCATGATCAAGCCCGCAGGCTTGACAGCA 2433
    |||||
QY 1910 GGGAGAAGCAGCCCGACCGTCTTC 1933
    |||
Db 2434 ATCACCAAAAATCTACCGTCTTC 2457
    |||

```

Search completed: December 12, 2002, 08:42:03
 job time : 400 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 12, 2002, 13:21:43 : Search time 74 Seconds

(without alignments)
4611.930 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646
Sequence: 1 MATTCPTADRGDAATDDP.....CPESGEHYETLLHPIQEYL 863

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Ygapext 0.5	
Xgapop 6.0 , Ygapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 355320 segs, 19730502 residues

Total number of hits satisfying chosen parameters: 710640

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US09976674/runat_04122002_162533_7241/app-query.fasta.1.1031
-DB=Published_Applications_NA -OFMT=fastap -SUFFIX=rp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=blcsunm2
-TRANS-human40.csl -LIST=45 -DCCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAULEN=200000000 -USER=US09976674@cgn2_1.1.33 -runat_04122002_162533_7241
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEM_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4646	100.0	2617	US-09-976-674-4	Sequence 4, Appl
2	4646	100.0	4219	US-09-976-674-28	Sequence 28, Appl
3	4646	100.0	4302	US-09-976-674-24	Sequence 24, Appl
4	4558.5	98.1	4180	US-09-976-674-36	Sequence 36, Appl

5	4558.5	98.1	4263	US-09-976-674-34	Sequence 34, Appl
6	4337	93.3	4076	US-09-976-674-32	Sequence 32, Appl
7	4337	93.3	4159	US-09-976-674-30	Sequence 30, Appl
8	4249.5	91.5	4037	US-09-976-674-40	Sequence 40, Appl
9	4249.5	91.5	4120	US-09-976-674-38	Sequence 38, Appl
10	2870	61.8	2671	US-09-976-674-12	Sequence 2, Appl
11	2850	61.3	4829	US-09-976-674-12	Sequence 12, Appl
12	2642.5	56.9	4685	US-09-976-674-22	Sequence 22, Appl
13	2608	56.1	4676	US-09-976-674-20	Sequence 20, Appl
14	2576	55.4	2411	US-09-976-674-26	Sequence 26, Appl
15	2548	54.8	2079	US-10-044-090-843	Sequence 843, Appl
16	2388.5	51.4	4523	US-09-976-674-8	Sequence 8, Appl
17	2204.5	47.4	4309	US-09-976-674-14	Sequence 14, Appl
18	647	13.9	1356	US-09-976-674-10	Sequence 10, Appl
19	498.5	10.7	4835	US-09-917-800A-1570	Sequence 1570, Ap
20	496	10.7	832	US-09-976-674-18	Sequence 18, Appl
21	487.5	10.5	3407	US-10-002-593-5	Sequence 5, Appl
22	487	10.5	281	US-09-967-550-987	Sequence 987, Appl
23	444	9.6	2583	US-09-976-674-6	Sequence 6, Appl
24	444	9.6	4541	US-09-976-674-42	Sequence 42, Appl
25	426	9.2	2814	US-09-954-531-367	Sequence 367, Ap
26	426	9.2	2814	US-09-962-832-108	Sequence 108, Appl
27	426	9.2	2814	US-09-954-456-1148	Sequence 1148, Ap
28	426	9.2	2815	US-09-965-606-1	Sequence 1, Appl
29	411.5	8.9	4496	US-09-976-674-44	Sequence 44, Appl
30	266.5	5.7	535	US-09-604-287A-428	Sequence 428, Appl
31	266.5	5.7	535	US-10-007-805-428	Sequence 428, Appl
32	249	5.4	1958	US-09-974-300-224	Sequence 224, Appl
33	216.5	4.7	2383	US-10-044-090-521	Sequence 521, Appl
34	193	4.2	497	US-10-046-935-2043	Sequence 2043, Ap
35	183	3.9	539	US-09-878-178-2043	Sequence 2043, Ap
36	183	3.9	539	US-09-878-178-2021	Sequence 2021, Ap
37	181	3.9	529	US-09-960-352-184	Sequence 184, Appl
38	181	3.9	529	US-09-976-674-16	Sequence 16, Appl
39	168.5	3.6	244	US-09-974-300-338	Sequence 338, Appl
40	166	3.6	620	US-09-966-803-1	Sequence 1, Appl
41	156.5	3.4	1797	US-09-728-445-436	Sequence 436, Appl
42	153	3.3	1869	US-09-934-223-3	Sequence 3, Appl
43	142	3.1	312	US-09-934-223-3	Sequence 3, Appl
44	118.5	2.6	2508	US-09-934-223-3	Sequence 1, Appl
45	118.5	2.6	4667	US-09-934-223-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-976-674-4
: Sequence 4, Application US/09976674
: Patent No. US20020115843A1
GENERAL INFORMATION:
: APPLICANT: Ol, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierre
: APPLICANT: Junien, Jean-Louis
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTV
: FILE REFERENCE: 70669
: CURRENT APPLICATION NUMBER: US/09/976,674
: PRIOR FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,117
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 4
: LENGTH: 2617
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-976-674-4

Alignment Scores:
Pred. No.: 0
Score: 4646.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Length: 2617
Matches: 863
Conservative: 0
Mismatches: 0

Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-976-674-3 (1-863) x US-09-976-674-4 (1-2617)

QY	1	MetIaIaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaIaIaThrAspPro	20
Db	11	ATGGCCACACACGGGGACCCCAAGCGCCGACGGAGGCGACACACCCCAAGATGACCG	70
QY	21	AlaIaIaArgPheGlnValGlnLysHisSerTPAspGlyLeuArgSerIleIleHisGly	40
Db	71	GGCCCCCGCTTCACAGTGTGCAGAAACACTCTGTGGAGAGGGCTCCGAGACATCATCCAGGC	130
QY	41	SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln	60
Db	131	AGCCGCAAGTACTGGGGCTCATTTGTCAMACAAAGGCCCCACACATTCAGATTGTGACAG	190
QY	61	LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrIleuGlyMetProTyrGly	80
Db	191	AAGAGGATGTGCTGTGGCCCACTCCACGGCTCTTACTACTGTGGAAATGGCAATATGGC	250
QY	81	SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu	100
Db	251	AGCGGAGAGAACTCCCTCCCTACTCTGAGATTTCCACAGAAAGTCCGGAAGAAGCTCTG	310
QY	101	LeuLeuLeuSerTyrPlyGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal	120
Db	311	CTGCTCCGTCTCTGGAAAGATGCTGATCATTTCCAGGCCACGCCCCACCATGTGGGTC	370
QY	121	TyrSerArgGluGluGluIleuLeuArgGluIaArgLysArgLysGlyValPheGlyIleThr	140
Db	371	TACTCTGGGGAGGAGAGCTGTGAGGAGCGGAAACGCTTGGGGTCTTCTGGCATATCAC	430
QY	141	SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	431	TCTACACACTCCACACGAGAGTGGGCTTCTCTCTCCAGGCCAGCAACAGCCTTTC	490
QY	161	HisArgArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
Db	491	CACGTGTGCGAGCGGGCAAGACGGCTTCATGTGTCTCCCTATGAACCGGTGGAAATC	550
QY	181	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	551	AAGACCCAGTCTAGAGGCCCGGATGTGACACCAAAATCTCCCTCCGACCTGCTTC	610
QY	201	PheSerPheIleAsnAsnSerAspLeuTyrPValAlaAsnIleGluThrGlyGluGluArg	220
Db	611	TTTCTCTTCATCATATACACGACGACTGTGGGTGGCCAAATCATGACAGCAGCAGGAGCGG	670
QY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerLysArg	240
Db	671	CGGCTGACCTTCTGCCACCAAGGTTTATCCAAATGCTTGATGATACCCCAAGTCTCGGGT	730
QY	241	ValAlaThrPheValIleGlnGlnGluIuPheAspArgPheThrGlyTyrTyrPrrPrrPrr	260
Db	731	GTGGCCACCTTCGTCATACAGAAAGATTTGACCGCTTCACTGGGTACGTGAGTGGTGGCC	790
QY	261	ThrAlaSerTrrPgluglySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal	280
Db	791	ACACCCCTTCGGGAAGGTTTCAGAGGGCTTCAGACGCTCGAATCTGTATAGGAAGTTC	850
QY	281	AspGluSerGluValGluValIleHisValProSerProAlaLeuGlnGluLysThr	300
Db	851	GATGATGCCAGTGGAGGTTCATTCACGTCCTCTCTCGCGCTTAAACAAAGGAAGACG	910
QY	301	AspSerTyrArgTyrProArgThrGlySerTyrAsnProLysIleAlaLeuLysLeuAla	320
Db	911	GACTCGATGTGATACCCACGACGACGACGAGAAGATCCCAAGTTGGCTTGAATGGCT	970
QY	321	GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln	340
Db	971	GAGTTCCAGACTACAGCCAGGGCAAGATTCGTCACCCAGAGAAAGAGCTGTGTGAG	1030

QY	341	ProPheSerSerLeuPheProLysValGluTyrGlnLeuAlaArgLysThrPheArgAsp	360
Db	1031	CCCTTCACACTCGCTGTTCCCGAAGGTGGATGACATCGCCAGGGCCGGGTGGACCCGGAGT	1090
QY	361	GlyLysTyrAlaThrPheLeuMetPheLeuAspArgProGlnGlnIleThrLeuGlnIleValLeu	380
Db	1091	GGGCAATACGGCTGGCCATGTTCTGTGAGCGGGGCCAGAGAGGTCTCCAGCTCTGCTCC	1150
QY	381	LeuProProAlaLeuPheIleProSerThrGlnAsnGluGlnIleArgLeuAlaSerAla	400
Db	1151	CTCCCCCGGGCCCTGTTCAITCCGAGCAGACAGAAATGAGAGCAGCGGTACGCTGTGCC	1210
QY	401	ArgAlaValProArgAsnValGlnProTyrValIleTyrGluGluValThrAsnValTrp	420
Db	1211	AGAGCTGTCCCCAGGAATGTTCACAGCCGTATGTGTGTACGAGAGAGGTACCAACGTCTGG	1270
QY	421	IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluLysAspLeuLys	440
Db	1271	ATCATATGTTCAATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGCAGCAGCTGTGC	1330
QY	441	PheLeuArgAlaAsnGluCysLysThrGlnPheCysHisLeuTyrTyrValIleThrAlaVal	460
Db	1331	TTTTCTCGCGCCCATGATGATGACAGACCGGCTTGTGCATTTGTACAAATCACCGCGCTT	1390
QY	461	LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluLysAspLeuPheLys	480
Db	1391	TTAAATATCCAGGGCTACGATGTGGAGTAGACCCCTTCAGCCCGGGGGAAGATGAATTTAAG	1450
QY	481	CysProIleLysGluGlnIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHis	500
Db	1451	TGCCCCATTAAGAAAGAAATGTGCTCTACACGCGGTAAATGGAGAGTTTGGCGAGGAC	1510
QY	501	GlySerLysIleTyrValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp	520
Db	1511	GGCTTCACAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCCACCAAGAC	1570
QY	521	ThrProLeuGlnHisLeuTyrValSerTyrGluAlaAlaGluLysIleValArg	540
Db	1571	ACGGCGGTGGAGCACACCTCTACGTGTGACGTATGAGCGGGCGGACAGATCGTACGC	1630
QY	541	LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheAl	560
Db	1631	CTCACACAGCCCGGCTTCTCCATAGCTGTCCATGAGCAGAACTCTGCACATGTTGCTGC	1690
QY	561	SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro	580
Db	1691	AGCCAGCTACAGACAGCGTACAGCCGCCCTGCGGTGACAGTCTACAGCTGACGGGCCCC	1750
QY	581	AspAspAspProLeuHisLysGlnProArgPheThrAlaSerMetMetGluAlaAlaSer	600
Db	1751	GAGCAGACACCCCTGTGCACAGACAGCCCGGCTTGTGGGTACAGATGATGGAGGCACCCAGC	1810
QY	601	CysProProAspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArg	620
Db	1811	TGCCCCCGGGATTAATGTTCTCTCCAGAGATCTTCCATTTCCACACCGCGCTCGGATGTGGG	1870
QY	621	LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal	640
Db	1871	CTCTACGGCATGATCTACAGGCCCCACAGCCTTGTGACGCCAGGAGAACAGCACCACCGCTC	1930
QY	641	LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys	660
Db	1931	CTCTTTTATATATGAGAGCCCCCAGGTGACAGCTGTGAATTAATCTCTTCAAGAGCATCAAG	1990
QY	661	TyrIleuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValValIleAspLysArg	680
Db	1991	TACTTGGGGGTCAACACACACAGGGCTCCCTGGGGCTACGCGCTGTGTGTGATGACGGCAGG	2050
QY	681	GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyLysIleVal	700
Db	2051	GGCTCTCTGTACGCGAGGGCTTCGGTTTGTGAAGGGGGCCCTTAATAAACCAATGTGGCCAGGTG	2110
QY	701	GluIleGlnAspGlnValGlnGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp	720

|||||
Db 2111 GAGATCGAGACAGGAGGAGGAGGAGTGCAGTCTGGCCGAGAAATATGAGCTTCATCGAC 2170
Qy 721 LeuSerArgValAlaIleHisGlyTrpSerTyrglygylPheLeuSerLeuMetGlyLeu 740
Db 2171 CTGAGCCAGATGGCCATCATGCTGCTGTCTACAGGGGGCTTCCTCTGCTCATGAGGGCTA 2230
Qy 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
Db 2231 ATCCAAAGCCCCAGGTGTTCAGAGTGGGACATCGGGGGTCCCGGTCAACCTCTGGATG 2290
Qy 761 AlaTrpAspThrGlyTyrglyThrGluArgTyrglyMetAspValProGluAsnGlnHisGly 780
Db 2291 GCCTACGACACAGGGGTACACTAGCGCTACATGAGCGTCCCTGAGAACACAGACAGCGC 2350
Qy 781 TyrglyAlaGlySerValAlaLeuHisValGlyLysLeuProAsnGluProAsnArgLeu 800
Db 2351 TATGAGGGGGTTCCTGCTGGCCCTGCACGCTGGAAGAGTCCCAATGAGCCCAACCGCTTG 2410
Qy 801 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
Db 2411 CTTATCTCCACAGGCTTCTCTGACGAAACGTCGACTTTTTCACACAACTTCTCGTC 2470
Qy 821 SerGlnLeuIleArgAlaGlyLysProTyrglyLeuGlnIleTyrglyProAsnGluArgHis 840
Db 2471 TCCCAACGATCCGAGGAGGAGAACTTACCAAGCTCCAGATCTACCCCAAGACAGACAC 2530
Qy 841 SerIleArgCysProGluSerGlyLysIleHisTyrglyValThrLeuLeuHisPheLeuGln 860
Db 2531 AGTATCTCCTGCTGCCAGTGGGCGAGCGACATGAAGTCACGTCTGCACTTCTACAG 2590
Qy 861 GluTyrlleu 863
Db 2591 GAATACCTC 2599

RESULT 2

US-09-976-674-28

Sequence 28, Application US/09976674

Patent No. US20020115843A1

GENERAL INFORMATION:

APPLICANT: Q1, Steve

APPLICANT: Akinsanya, Karen

APPLICANT: Riviere, Pierre

APPLICANT: Junien, Jean-Louis

TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLY

FILE REFERENCE: 70669

CURRENT APPLICATION NUMBER: US/09/976,674

PRIORITY FILING DATE: 2001-10-12

PRIOR APPLICATION NUMBER: US 60/240,117

PRIORITY FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 61

SOFTWARE: PatentIn version 3.1

SEQ ID NO 28

LENGTH: 4219

TYPE: DNA

ORGANISM: Homo sapiens

US-09-976-674-28

Alignment Scores:

Pred. No.:	0	Length:	4219
Score:	4646.00	Matches:	863
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-976-674-3 (1-863) x US-09-976-674-28 (1-4219)

Qy 1 MetaIaIaThThGlyThrProThrAlaAspArgGlyAspAlaAlaIaIaThrAspPro 20
|||||
Db 367 ATGGCACCACCGGAGCCCAAGCGGCGACCGAGCGAGCGAGCGCCCAACATGACCGG 426
Qy 21 AlaIaIaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40

|||||
Db 427 GCCGCCCGCTTCAGATGAGAACACATCTGTGAGCGGGGCTCCGAGCATATCCACGCGC 486
Qy 41 SerArgLysTyrglySerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCCCAAGTACTCGGGCTCATTTGTCAACAAGGGCGCCCAACGACTTCAGATTGTGACG 546
Qy 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrglyLeuGlyMetProTrpGly 80
Db 547 AAGACGAGTACTGTGGGCCCCACTCCACCGCTCTACTACTCTGTGGAAATGCCATATGCG 606
Qy 81 SerArgLysAsnSerLeuLeuTyrglySerGluIleProLysLysValArgLysGluAlaLeu 100
Db 607 AGCCGAGAAATCCTCCTCCTCTACTAGATTCACCAAGAGCTCCGAAAGAGGCTCTGTG 666
Qy 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 667 CTGCTCCCTGCTCTGAGACAGATGCTGATCATTTCCAGGGCCACGCCCAACCATGGGGTTC 726
Qy 121 TyrglySerGluGluGluLeuLeuArgGluArgLysArgGlyValPheGlyIleThr 140
Db 727 TACTCTCGGAGAGAGAGCTGCTGAGGGAGCGGAAACGCTGGGGTCTTGGCATCACCC 786
Qy 141 SerTyrglyAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 787 TCCCTACGACTTCCACAGCAGAGAGTGGCTCTTCTCTTCCAGACGACGACGCTCTTC 846
Qy 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 847 CACTCGCGGAGCGGCGGCAAGAGCGCTTCATGTGTCCCTCATGAAACCGCTGGAATC 906
Qy 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
Db 907 AAGACCCAGTGTCTCAGGGGCCCGGATGAGACCCCAAAATTCGCCCTGCCACCTCGCTTC 966
Qy 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 220
Db 967 TTCTCTCTATCATATATACAGCAGCACTGTGTGGCCAAACATGAGACAGACGAGACGCG 1026
Qy 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 1027 CGGCTGACCTTCTGCGACCAAGGTTATCCATGTCTGATGAGACCCCAAGCTCGGGGT 1086
Qy 241 ValAlaIaIaThrPheValIleGlnGluPheAspArgPheThrGlyTrpTrpPro 260
Db 1087 GTGGCCACCTTCTGCTATACAGGAAGATTCGACCCCTTCTACTGAGTGGGTGCGCC 1146
Qy 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrglyGluVal 280
Db 1147 ACAGCCTCTGGGAGAGGTTCAAGAGGCGCTCAAGAGCGCTCGAATCTGTATAGAGAACTC 1206
Qy 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
Db 1207 GATGAGTCCGAGGTGAGGTCAATTCACGCTCCCTCTCTGCTGAGAAAGAAAGACG 1266
Qy 301 AspSerTyrglyArgTyrglyProArgThrGlySerLysAsnProLysIleAlaLeuLysVal 320
Db 1267 GACTCGTATCGGTACCCCGAGGAGCGACAGCAAGATCCCAAGATTGCTTGAACCTGGCT 1326
Qy 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluGlyLeuValGln 340
Db 1327 GAGTTCACAGACTGACAGCGGCAAGATCGTCTGACCCACGAGAAAGAGTGTGGCAG 1386
Qy 341 ProPheSerSerLeuPheProLysValGluTyrlleAlaArgAlaGlyTrpThrArgAsp 360
Db 1387 CCTTCACGCTCGCTGTCCCGAAGGTGAGTACATGCGCAGGGCGGAGCGGCGGAT 1446
Qy 361 GlyLysTyrglyAlaThrAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnValLeu 380
Db 1447 GGCAAATATACGCTGGGCGCATGTTCTGAGCGCGGCCCAACAGAGGCTCCAGCTCGCTTC 1506
Qy 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAla 400
|||||

Db 1507 CTCCCCCGCCCTGTCATCCGACAGATGAGGACGCGGCTCTGCC 1566
QY 401 ArgAlaValProArgAsnValGlnProTyrValTyrGluGluValThrAsnValTyr 420
Db 1567 AGAGCTGTCCCGAGGAAATGTCACGCCGATGCGTACGAGGAGGTCCACCAACGCTCGG 1626
QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGluLeuGly 440
Db 1627 ATCATATGTCATGACATCTCTATCCCTCCCAATCAGAGGAGAGGACGACCTCTGC 1686
QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
Db 1687 TTTTCTCCGCCCAATGATGCAAGACGCGCTTGCCATTGTACAAAGTCAACGCGGTT 1746
QY 461 LeuLysSerGlnGlyTyrAspTyrSerSerGluProPheSerProGluGluAspGluPheLys 480
Db 1747 TTTAAATCCCGAGGCTTCAGATGAGTGCCTTCAGCCCGGGGAGAAATGATTTTAAG 1806
QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyLutRgLuValLeuAlaArgHis 500
Db 1807 TGCCCCATTAAAGAAAGATTGCTCTGCACAGCGGTGATGGAGGTTTGGCAGCGCAC 1866
QY 501 GlySerLysIleTyrValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
Db 1867 GGTCTCCAGATCTGGGTCAATGAGGAGACCAAGCTGGGTGATCTTCCAGGGACCAAGGAC 1926
QY 521 ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaAlaGluIleValArg 540
Db 1927 AGCGCGGTGGAGACCAACCTCTACGTGTCATGAGCGCGCGGAGATGTTGTCACGC 1986
QY 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
Db 1987 CTCACACGCGCGGCTTCTCCATAGCTGCTCCATGAGCCGAACCTTCGACATCTTCGTC 2046
QY 561 SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro 580
Db 2047 AGCCACTACACACGCGTGCAGACGCCGCCCTGCTGCACGCTTCACAACTGAGGCGCCC 2106
QY 581 AspAspAspProLeuHisLysGlnProArgPheTyrPheTyrPheTyrPheTyrPheTyr 600
Db 2107 GACGACGACCCCTGCACAAACGACCCCGCTTCTGGGTACCATGATGAGGACGACGAC 2166
QY 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
Db 2167 TGCCCCCGGGTATGTTCTCCACAGATCTTCATTCACACGCGCTCGATGTGCGG 2226
QY 621 LeuTyrGluMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
Db 2227 CTCTACGGCATGATCTACAAAGCCCGCGCTTGCGACGCCAGGAAAGACACCCACCGTC 2286
QY 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
Db 2287 CTCTTTTATATGAGGCGCCCGACGTCAGCTGTGTAATACCTCTTCAAAAGCATCAG 2346
QY 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 680
Db 2347 TACTTGGGGTCAACACACTGCGCTCCCTGGGCTACCGCGGTGTTGATGAGCGCAGG 2406
QY 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyIleVal 700
Db 2407 GGCTCTCTGTCAGCGAGGCTTCGCTTGAGAGGCGCCCTGAAAACCAAAATGGGCGAGGTG 2466
QY 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
Db 2467 GAGATGACAGGACGAGTGGAGGCGCTGACAGTTCGTGGCCGAGAGTATGGCTTATGAGAC 2526
QY 721 LeuSerArgValAlaIleHisGlyTyrPheTyrGlyLysPheLeuSerLeuMetGlyLeu 740
Db 2527 CTGAGCCGATGTCATTCATGCTGCTACAGGGGCTTCCTCTGCTCATATGGGCTA 2586
QY 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTyrMet 760
Db 2587 ATCCACAAGCCCGAGGTTCAGAGTTCGATCCGCGGCTGCCCGGACCGTCTGGATG 2646

QY 761 AlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly 780
Db 2647 GCTTACGACACAGGGGTACACTGAGCGCTACATGAGACGCTCCCTGAGAACACACACGCGC 2706
QY 781 TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu 800
Db 2707 TATAGGCGGGTTCGCGGCGCTGACGTGGAGAAAGTGGCCCAATGAGCCCAACCGGTTG 2766
QY 801 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
Db 2767 CTTATCTCCACGCGCTTCCTGAGCGAAACGTCGACTTTTCCACAAACTCTCTGTC 2826
QY 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
Db 2827 TCCCACTGATCCGAGCAGGAGGAACCTTACAGCTCCAGATCTACCCCAACGAGAGACAC 2886
QY 841 SerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGln 860
Db 2887 AGTATTCGCTGCGCCGAGTGGGCGGACACTATGAACTGACGTTGTGCACCTTCTACAG 2946
QY 861 GluTyrLeu 863
Db 2947 GAATACCTC 2955

RESULT 3
US-09-976-674-24
; Sequence 24, Application US/09976674
; Patent No. US20020115843A1
; GENERAL INFORMATION:
; APPLICANT: Qi, Steve
; APPLICANT: Akinsanya, Karen
; APPLICANT: Riviere, Pierre
; APPLICANT: Junien, Jean-Louis
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DEPTV
; FILE REFERENCE: 70669
; CURRENT APPLICATION NUMBER: US/09/976, 674
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,117
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 4302
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-976-674-24

Alignment Scores:
Pred. No.: 0 Length: 4302
Score: 4646.00 Matches: 863
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-976-674-3 (1-863) x US-09-976-674-24 (1-4302)

QY 1 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspValAlaIleThrAspAspPro 20
Db 367 ATGGCCACCAACCGGACCCCAACAGCGCGAGCGGACGCGGCGGACGAGATGAGCCG 426
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTyrAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCGCGCGGCTTCAGAGTGCAGAAAGCACTGCGGAGCGGCTCCGAGACATTCACAGGC 486
QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCGCAAGTACGCGGCTCATTTGTCACAAAGCGCCCAACGACTTCCAGTTTGTGACAG 546
QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGCGATGAGTCTGGGCGCCCACTCCACCGCTTACTACTGAGGAAATGCCATTATGGC 606

QY	81	SerATrgLuanSerLeuLeuTYrSerGluIleProLysValArgLysGluAlaLeu	100
Db	607	AGCGAAGAAACTCCCTCTACTCTGAGATTCCCAAGAAAGTCCGGAAGAGACTCTG	666
QY	101	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal	120
Db	667	CTGCTCTGCTCTGGAAAGCATGGATGATCATTTTCCAGGCCACGCCCTCCACTGGGGTTC	726
QY	121	TyrSerATrgLugLugLueLeuATrgGluValArgLysArgLeuGlyValPheGlyValThr	140
Db	727	TACTCTGGGAGAGAGAGCTGCTGAGGAGACCGAAACGCTTGGGGTCTTGGCATCAC	786
QY	141	SerTYrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
Db	787	TCTTACAGACTTCCACACCGAGAGTGGCCCTTCTCTTCCAGGCCAGACAGCCTCTTC	846
QY	161	HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
Db	847	CACGCGCGACGCGCGGCAAGAAAGGCTTCATGTGTCTCCCTATGAAACCCCTGGAATC	906
QY	181	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe	200
Db	907	AAGACCCAGTGTCTCAGGGCCCCGGATGGACCCCAAAATGTGCGCTGGACCTGTCTTC	966
QY	201	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluValArg	220
Db	967	TTTCTCTTATCATATACAGCAGACTGTGGTGGCCAACTGAGACAGCGGAGAGAGGG	1026
QY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly	240
Db	1027	CGGTGACCTTCTGCGCCACCAAGGTTTATCCATGTCTCGGATGAGACCCCAAGCTGCGGGT	1086
QY	241	ValAlaThrPheValIleGlnGlnGluIupheAspArgPheThrGlyTyrTrpTrpCysPro	260
Db	1087	GTGGCCACTTTCGTATACAGAGAAAGATTGACCGCTTACTGGGTACTGGGTGGCCCC	1146
QY	261	ThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGluVal	280
Db	1147	ACAGCCTCTCGGGAAGGTTCAGAGGGCCCTCAAGACCTCGAATCTGTATGAGGAAGTC	1206
QY	281	AspGluSerGluValGluValIleHisValProSerProAlaLeuGlnGluArgLysThr	300
Db	1207	GATAGATCCGAGGGGAGGATCATTCACGTCCTCTCTCGGCTAGAGAAAGGACAGC	1266
QY	301	AspSerTrpArgTyrProArgThrGlnLysTyrAsnProLysIleAlaLeuLysLeuAla	320
Db	1267	GACTCGTATCGGTACCCCAAGACAGGACAGCAAAATCCCAAGATTGCTTGAACCTGGCT	1326
QY	321	GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln	340
Db	1327	GAGTTCAGACTGACACCGAGGGCAGAGATCGTCTGCACCCAGAGAAAGAGACTGGTGAG	1386
QY	341	ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrTrpArgAsp	360
Db	1387	CCCTTCAGCTGTGCTTCCGGAAGGTGGATCATGCGCCAGGGCGGGTGGACCCGGGAT	1446
QY	361	GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu	380
Db	1447	GGCAAAATPACGCTGGGCGCATGTTCTGGACCGGCCCCCAACAATGGCTCTCAGCTGCTCTC	1506
QY	381	LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGluArgLeuAlaSerAla	400
Db	1507	CTCCCCCGGCGCTTTCATCCCGAGACAGAAATGAGAGAGACAGCGCTAGCCTTGCC	1566
QY	401	ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGluValThrAsnValTrp	420
Db	1567	AGAGCTGTCCCCAGAGATGTCCACGCGTATGTGTGTACGAGAGAGATACCAACGCTGCG	1626
QY	421	IleAsnValHisAspIlePheTyrTrpPheProGlnSerGlnGlyGluAspGluLeuCys	440
Db	1627	ATCATATGTTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGACAGCTGTGC	1686

QY	441	PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal	460
Db	1687	TTTTCTCCGGCGCATGAAATCAAGACCGGGCTTCCTGGCATTTTGTACAAAGTACACCGCCCTT	1746
QY	461	LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyLysAsnGluPheLys	480
Db	1747	TTTAAATCTCCAGGGCTACGATGTGAAGTGAAGCCCTTCAAGCCCGGGGGAAGATGAATTAAAG	1806
QY	481	CysProIleLysGlnGluIleAlaLeuThrSerGlyLysGluThrGluValIleuAlaArgHis	500
Db	1807	TGCCCATTTAAGGAAGAGATGTGCTGACACAGGGGTGAATGGAGAGTTTGGCAGAGCAC	1866
QY	501	GlySerLysIleTrpValAsnGluLutPheLysLeuValTyrPheGlnGlyThrLysAsp	520
Db	1867	GGCTCCAAAGATCTGGGTCCATGTAGAGACCAACAGCTGGTACTTCCAGGCGACCAAGAC	1926
QY	521	ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaAlaGlyLutIleValArg	540
Db	1927	ACGGCGCTGGAGACCCACCTCTACGTGGTCACTAAGAGCGGGCGGAGATCTGTACGC	1986
QY	541	LeuThrTrpProGlyPheSerHisSerCysSerMetSerGluAsnPheAspPheVal	560
Db	1987	CTCACACAGCCCGGGCTTCTCCCATAGCTGCTCATATGACCAAACTTGACATGTTCGTC	2046
QY	561	SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro	580
Db	2047	AGCCACTACACACACGTCGACACGCGCCCTCGCTGCATCTACAGCTGAGCGGGCCC	2106
QY	581	AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetIleGluAlaAlaSer	600
Db	2107	GACGACGACCCCTCGACCAAGACGACCCCGCTTGTGGCTAGATGATGAGACACCGACGC	2166
QY	601	CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg	620
Db	2167	TGCCCGCGGATTTATGTTCTCTCCAGAGATCTTCATTCCACACCGCTCGGATGGCG	2226
QY	621	LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal	640
Db	2227	CTTACGCGCATGATCTCAAGACCCCAAGCGCTTGCACGCCAGGAAAGACACCCCAACGTC	2286
QY	641	LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLys	660
Db	2287	CTCTTGTATATGAGAGGCCCAAGGAGGCACACTGTGTAAATCTCTTCAAAAGCATCAAG	2346
QY	661	TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg	680
Db	2347	TACTTCGCGCTCAACACACTGGCTCCCTGGCTGACCGCTGTGTGATTGACGGCACGG	2406
QY	681	GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyLysVal	700
Db	2407	GGCTCTCTGACGAGGGCTTCGGTTCTGAAAGGGGGCCCTGAATAAACAATGGCCAGGTG	2466
QY	701	GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp	720
Db	2467	GAGATCGAGGACCAAGGTGGAGGGCCCTGTCACTTGTCGGCGAAGAAGATGGCTTCATCCAC	2526
QY	721	LeuSerArgValAlaIleHisGlyTyrPseTyrGlyGlyPheLeuSerLeuMetLysLeu	740
Db	2527	CTGAGCCGAGTGGCATTCATGGCTGGTCTCTACGGGGGGCTTCCTCTCGCTCATGGGGCTA	2586
QY	741	IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet	760
Db	2587	ATCCACAAGCCCCAGGGTTCCAAGGTGGCATGGGGGGTCCCGGTCACCGTCGATG	2646
QY	761	AlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly	780
Db	2647	GGCTACGACACAGGGTCACTGTAGCGGTACATGAGAGTCCCTGAGAACACACAGCACCGC	2706
QY	781	TyrGluLysGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu	800
Db	2707	TATGAGCGGGGTCTCGTGGCCCTGTCAAGTGAAGAACTGCCAATAAGGCCCAACCGCTTG	2766
QY	801	LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal	820

|||||
Db 2767 CTTATCCCGCAGGCTTCTGAGCAAAACGAGCACTTTTCCACAAACCTTCTCTG 2826
Qy 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
Db 2827 TCCCACTGACATCCGAGCGAGGAAACCTTACCAAGCTCCAGATCTTACCCCAAGAGAGACAC 2886
Qy 841 SerIleArgCysProGluSerGlyGlnHisTyrGluValThrLeuLeuHisPheLeuGln 860
Db 2887 AGATATCCCTGCCCCGAGCTGGGCGAGCATATGAATGACATGCTGCTGCTCTCTACAG 2946
Qy 861 GluTyrIleu 863
Db 2947 GAATACCTC 2955
RESULT 4
US-09-976-674-36
Sequence 36, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Q1, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPPIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976, 674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240, 117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36
LENGTH: 4180
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-36
Alignment Scores:
Pred. No.: 0 Length: 4180
Score: 4558.50 Matches: 850
Percent Similarity: 98.49% Conservative: 0
Best Local Similarity: 98.49% Mismatches: 0
Query Match: 98.12% Indels: 13
DB: Gaps: 1
US-09-976-674-3 (1-863) x US-09-976-674-36 (1-4180)
Qy 1 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaIleThrAspAspPro 20
Db 367 ATGGCCACACCGGAGCCGCAACGCGCAGCGAGCGAGCGCGCCACAGATGACCCG 426
Qy 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
Db 427 GCCGCCCGCTTCCAGGTGTCAGAACACACTCTGTGGACGGGCTCCGAGCATATCCACGGC 486
Qy 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 487 AGCGCAAGTACTCGGGCTCATTTGTCAACAAGCGGCCCGCCAGCACTTCAGATTGTGCAG 546
Qy 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
Db 547 AAGACGATGTGGTGGGCCCACTCCACCGCTTACTACCTGGGGAATGCCATATGGC 606
Qy 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
Db 607 AGCCGAGAGAACTCCCTCTCTACTCTGAGATTCCCAAGAAAGGTCCGAAAGAGGCTTG 666
Qy 101 LeuLeuLeuSerTyrPlySglnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 667 CTGCTCCGTCTCGAAGACATGCTGATCATTTCCAGGCGACGCCCCACCATGAGGGCTC 726
Qy 121 TyrSerArgGluGluGluIleLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 140

|||||
Db 727 TACTCTCCGGAGGAGGAGCACTCTGAGCGAGGAAACGCTGGGGCTTCCGATCAC 786
Qy 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 787 TCTTACGACTTCCACAGGAGAGTGGCTCTTCTCTTCCAGGCCAGCAAGCTCTTCC 846
Qy 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 847 CACTGGCCCGCAGCGGGGCAAGAGCGCTTCATGTGGTCCCTATGAAACCGCTGGAATC 906
Qy 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
Db 907 AAGACCCAGTCTCAGGGCCCCGGATGAGCCCAAAATCTGCCCTGCCAGCTCCCTTC 966
Qy 201 PheSerPheIleAsnAsnSerAspLeuThrValAlaAsnIleGluThrGlyGluGluArg 220
Db 967 TTCTCTTCAATCAATACAGGAGCACTGTGGGTGGCCAAATCGACATCGACAGGAGCGG 1026
Qy 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 1027 CGGCTGACCTTCTGCGACCAAGGTTTATCCATATGCTCGATGAGCCCAAGTCTGGGGT 1086
Qy 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCysPro 260
Db 1087 GTGGCACCTTCGTCAATACAGGAAGATTCGACCGCTTCACTGGGTACTGGTGGCCCC 1146
Qy 261 ThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluVal 280
Db 1147 ACAGCTCTCTGGAGAGGTTCCAGAGGCTCTCAAGACGCTGCAAAATCTGTATGAGAACTC 1206
Qy 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
Db 1207 GATGAGTCCGAGGTGGAGTATTCACGCTCCCTCTCTCGGCTAGAAAGAAAGAGACG 1266
Qy 301 AsperTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
Db 1267 GACTCGTATCGGTACCCCGAGCAGGCGCAAGAAATCCCAAGATGCTCTGAAACTGGCT 1326
Qy 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluValGln 340
Db 1327 GAGTTCCAGACTGACAGCGACGCAAGATCTCTCGACCCAGAGAGAGCACTGGTCAG 1386
Qy 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrTrpArgAsp 360
Db 1387 CCGTTCAAGCTCGTGTGCCAAGGTGGATACATCGCCAGGCGGGGTGACCCGGGAT 1446
Qy 361 GlyLysTyrAlaIleThrPheLeuAspArgProGlnGlnThrPheLeuValLeu 380
Db 1447 GGCAAATACGCTGGCCATGTTCTGTGACCGGCCCGCCAGCGTGGCTTCACCTGCTCTC 1506
Qy 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAla 400
Db 1507 CTCCCCCGGCGCTGTTCTATCCGACGACAGAGATGAGAGCAAGGCTGCTCTGCC 1566
Qy 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrp 420
Db 1567 AGAGCTGTCCCGAGAAATGTCCAGCCCGTATGTGTGACGAGAGGTCCACCAAGTCTCG 1626
Qy 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspGluLeuCys 440
Db 1627 ATCAATGTTCATGACATCTTATCTCTCCCTCCCAATCAGAGGAGAGAGCACTCTGCG 1686
Qy 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
Db 1687 TTTCTCCGCGCAATGAATGAAAGCCGCTTCTGCTTCTGTAACAAAGTACCGCGCTT 1746
Qy 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLys 480
Db 1747 TTAATAATCCAGGGCTAGATGGAGTGAGCCCTTCAAGCCCGGAGAAAGATGAAATTAAG 1806
Qy 481 CysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHis 500

Db 1807 TGCCCATTAAGAGAGATTGCTCTGACACAGCGTGAATGGAGTTTGGCGAGCAC 1866
 QY 501 GlySerLysIleTrpValAsnGluGlnTrpLysLeuValTyrPheGlnGlyThrLysAsp 520
 Db 1867 GGCTCCAG-----GGCACCAAGGAC 1887
 QY 521 ThrProLeuGluHisIleLeuTyrValValSerTyrGluAlaIleGluIleValArg 540
 Db 1888 ACGCGCTGGAGACACACCTGACCTGACCTGATGAGGGCGCGGAGATGCTAGCC 1947
 QY 541 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 Db 1948 CTCACACAGCCCGCTCTCCATAGCTGCTCCATGAGCAGCAACCTTGACATGTTCCGC 2007
 QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
 Db 2008 AGCCACTACAGACGCTGAGCAGCGCGCCCTGCGTGCAGCTTACAGCTGAGCGGCCCC 2067
 QY 581 AspAspAspProLeuHisIleGlnProArgPheTrpAlaSerMetMetGluAlaIleAsp 600
 Db 2068 GACGACGACCCCTGACACAGACGCCGCTTCTGGGCTGACATGATGAGGACAGCCAGC 2127
 QY 601 CysProProAspTrpValProProGluIlePheHisPheHisThrArgSerAspValArg 620
 Db 2128 TGCCCCCGGATTATGTTCTCTCCAGAGATCTTCATTCCACACCGCGCTCGATGTCGG 2187
 QY 621 LeuTyrGluMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
 Db 2188 CTCATGGCATGATCTACAGACCCACGCGCTTGACGCCAGGAGAAAGACCCACCGCTC 2247
 QY 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
 Db 2248 CTCCTTGTATATGAGAGGCCCGCCAGCTGCGTGAATAACTCTCTCAAGCATCAAG 2307
 QY 661 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArg 680
 Db 2308 TACTTGGCGCTCACACACTGGCTCCCTCGGCGCTAGCGCGCTGTTGATGACGAGCAG 2367
 QY 681 GlySerCysGlnArgGlyLeuArgPheGluGluAlaLeuLysAsnGlnMetGlyIleVal 700
 Db 2368 GGCTCTCTACAGCGAGGCTTGGTTCGAAAGGGGCCCTGAAACCAATGGGCGCAGGG 2427
 QY 701 GluIleGluAspGlnValGluGluLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
 Db 2428 GAGATCGAGAGACACAGCTGAGAGGCCCTGCAGTTCGTGGCCGAGAAATGATGCTTCATCGAC 2487
 QY 721 LeuSerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 740
 Db 2488 CTGACCGCAGTGGCCATCCATGGCTGCTGCTGACGGGGCTTCCTCGCTCATGGGGCTA 2547
 QY 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
 Db 2548 ATCCCAAGCCCCAGGTGTCAAGGTGGCATCGCGGGTCCCGCTCACCTCTGGAGG 2607
 QY 761 AlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly 780
 Db 2608 GCGTCAGACACAGGGGTACACTAGCGGCTACATGAGAGCTCCCTGAGAACACACAGACAGCC 2667
 QY 781 TyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeu 800
 Db 2668 TATGAGGGGGGTCCGTGGCCCTGCACGTCGAGAGCTGCCAATGAGCCCAACGGCTTG 2727
 QY 801 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
 Db 2728 CTTATCTCCACGCGCTCTCTGAGCAAAACGTCATTCTTCCACAAATCTCTCCGCTC 2787
 QY 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
 Db 2788 TCCCAACATGATCCGAGACGAGGAAACCTTACAGCTCCACATCTACCCCAACAGAGACAC 2847
 QY 841 SerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGln 860
 Db 2848 AGTATTCGCTGCCGCGAGTCGGGCGAGCACTATGAAGTCACTGCTGCACTTCTACAG 2907

QY 861 GluTyrLeu 863
 Db 2908 GAATACCTC 2916
 RESULT 5
 US-09-976-674-34
 ; Sequence 34, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akinsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Junien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
 ; FIDE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976, 674
 ; PRIOR FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/240,117
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 34
 ; LENGTH: 4263
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-976-674-34
 Alignment Scores:
 Pred. No.: 0 Length: 4263
 Score: 4558.50 Matches: 850
 Percent Similarity: 98.49% Conservative: 0
 Best Local Similarity: 98.49% Mismatches: 0
 Query Match: 98.12% Indels: 13
 Gaps: 1
 US-09-976-674-3 (1-863) x US-09-976-674-34 (1-4263)
 QY 1 MetalAlaThrThrGlyTyrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 20
 Db 367 ATGGCCACACACCGGAGACCCCAACGCGCGACGAGCGAGCGCCGACAGATGACCCG 426
 QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
 Db 427 GCCGCCCGCTTCAGGTGCAGAAACACTGCTGGAGAGGGCTCCGAGCATCAACAGCG 486
 QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
 Db 487 ACCCGCAAGTACTCGGGCTCATTTGTCAACAAAGGGCCCAAGACTCCAGTTGTGGAG 546
 QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
 Db 547 AAGACGAGTACTGCGGCCCACTCCACGCTCTACTACTGGAATGCAATATGCG 606
 QY 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
 Db 607 ACCCGAGAACTCCCTCTCTACTCTGAGATTCACAAAGAGTCCGGAAGAGAGGCTCG 666
 QY 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
 Db 667 CTGCTCTCTCTCTGAGAGAGATGCTGATCTTTCACAGCCACAGCCCAACATGGGGTC 726
 QY 121 TyrSerArgGluGluGluLeuArgGluArgGlyValArgLeuGlyValPheGlyIleThr 140
 Db 727 TACTCTCGGAGAGAGAGTCTGAGGGAGCGGAAACGCTGCGGTCTTCCGCAATCAC 786
 QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
 Db 787 TCTTACGACTTCACAGCAGAGTGGCTCTTCTCTTCCAGGACGACAGAGCTTTC 846
 QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
 Db 847 CACTGCCGCGAGCGGGGAGAAAGCGCTTCATGATGATGCTCCATGAACCGCTGGAATTC 906

Db	2107	GAGNCGACCCCGTCGACAGCAGACGCCCGCTTCTGGGCTAGCATGATGAGAGCAGCCAGC	2168
QY	601	CysProProAspIyrValProProGluIlePheHisPheHisThrArgSerAspValArg	620
Db	2167	TGCCCCCGGGATTATGTCTCTCCAGAGATCTTCCATTCTACACCCGCGTGGATGTGGG	2226
QY	621	LeuIyrGlyMetIleIyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal	640
Db	2227	CTTACGCGCATGATCTACAGGCCCGACGCTTGCAAGCCAGGGAGAGACACCCACGCTC	2286
QY	641	LeuPheValIyrGlyIyrProGlnValGlnLeuValAsnAsnSerPheLysGlyLleLys	660
Db	2287	CTCTTTGATATGAGAGGCCCCCGACGTGAGCTGTGAATAATCTCTTCAAGAGCATCAAG	2346
QY	661	TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspIyrArg	680
Db	2347	TACTTGGGGCTCAACACATGGCTCCCTGGGTACGGCCGTGGTGTGAATTACGGCAGG	2406
QY	681	GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal	700
Db	2407	GGCTCTGTGTGAGCAGAGGGCTTTCGGTTTGAAGGGGGCCCTGAATAACCAATGGCGCAGGTG	2466
QY	701	GluIleGlnsArgGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp	720
Db	2467	GAGATTCAGAGCCCGAGGGGAGGGCCGTGCAAGTTCGTGGCGAACAATATGCTTCACTCAC	2526
QY	721	LeuSerArgValAlaIleHisGlyTyrPseTyrGlyIyrPheLeuSerLeuMetGlyLeu	740
Db	2527	CTGAGCCGAGTGGCATTCATCGCTGTGCTCTACGGGGGCTTCTCTCTCATGGGGCTA	2586
QY	741	IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTyrPme	760
Db	2587	ATCCACAGAGCCCGAGGTGTCAAG-----	2611
QY	760	TAlaTyrAspPheIcyIyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGln	780
Db	2611	-----	2611
QY	780	YTrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLe	800
Db	2612	-----CCACACCGCTT	2622
QY	800	WLeuIleLeuHisGlyPheLeuAspGlnsValHisPhePheHisThrAsnPheLeuVal	820
Db	2623	GCTTATCTCTCCAGCGCTTCTGTGACAGAAAACGGTTCCTTTTCCACAAAACCTTCTCGT	2682
QY	820	IserGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis	840
Db	2683	CTCCCACTGATCGACACAGGAACCTTATACAGGCTCCAGATCTACCCCAACAGAGAGCA	2742
QY	840	SserIleArgCysProGluSerGlyLunHisTyrGluValThrLeuLeuHisPheLeuGln	860
Db	2743	CAGATTTCGTCGCCCCGACAGTCGGGCGAGCAGCATATGAAGTACAGTGTGCTGCACTTCTACA	2802
QY	860	ngIuTyrLeu 863	
Db	2803	GGATFACCTC 2812	
RESULT 7			
; Sequence 30, Application US/09976674			
; Patent No. US20020115843A1			
; GENERAL INFORMATION:			
; APPLICANT: Qi, Steve			
; APPLICANT: Akinsanya, Karen			
; APPLICANT: Riviere, PIERRE			
; APPLICANT: Junien, Jean-Louis			
; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV			
; FILE REFERENCE: 70669			
; CURRENT APPLICATION NUMBER: US/09/976,674			
; PRIOR FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: US 60/240,117			
; PRIOR FILING DATE: 2000-10-12			

;	NUMBER OF SEQ ID NOS:	61	
;	SOFTWARE:	Patentin version 3.1	
;	SEQ ID NO	30	
;	LENGTH:	4159	
;	TYPE:	DNA	
;	ORGANISM:	Homo sapiens	
;	US-09-976-674-30		
Alignment Scores:			
Pred. No.:	0	Length:	4159
Score:	4337.00	Matches:	814
Percent Similarity:	94.33%	Conservative:	1
Best Local Similarity:	94.21%	Mismatches:	0
Query Match:	93.35%	Indels:	49
DB:	10	Gaps:	1
US-09-976-674-3 (1-863) x US-09-976-674-30 (1-4159)			
QY	1	MetaIaThrPrbGlyThrProThrAlaAspArgGlyAspAlaAlaIaThrAspAspPro	20
DB	367	ATGGCACCACCGGGGACCCACAGCGCGAGCGAGGAGCGACGCCGCCACATACACCG	426
QY	21	AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly	40
DB	427	GCCGCCCTTCACAGTCGACGAAGCACTCGCGGAGCGGCTCGGAGCATCCACCGGC	486
QY	41	SerArgLysTrpSerGlyLeuLysIleValAsnLysAlaProHisAspPheGlnPheValGln	60
DB	487	AGCCGCAGTACTCGGGCTCTATTGTCACAGAGGCGCCCGACACTTCAGATTGTGCAG	546
QY	61	LysThrAspGluSerGlyProHisSerHisArgLeuTyTrpLeuGlyMetProTyrGly	80
DB	547	AAGACGATGATGTGGGGCCCACTCCACCGCTCTACTCGGAAATGCCATATATGCG	606
QY	81	SerArgGluAsnSerIleLeuTyTrpSerGluIleProLysLysValArgLysGluAlaLeu	100
DB	607	AGCCGAGAACTCCCTCTACTCGAATATCCCAAGAGTCCGGAAGAGGCTCTG	666
QY	101	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaIaThrProHisHisGlyVal	120
DB	667	CTGCTCCGTCCTCGAAGCAGATGTCGAGATCATTTCCAGGCCACGGCCACCATGGGTC	726
QY	121	TyrSerArgGluGluGluLeuLeuAlaArgGluArgLysArgLysGlyValPheGlyIleThr	140
DB	727	TACTCTCGGGAGGAGGCTGTAGAGGACCGGAAACCGCTGGGGCTCTCCGCAATCAC	786
QY	141	SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe	160
DB	787	TCTTACGACTTCCACACGAGAGTGGGCTTCTTCTTCAGGCCAGCAACGCTTTC	846
QY	161	HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle	180
DB	847	CACTCGCCGCGACGGCGGCAAGAACGGCTTCATGTCTCCCATATAAACCGCTGGAATTC	906
QY	181	LysThrGlnCysSerGlyProArgMetLeuAspProLysIleCysProAlaAspProAlaPhe	200
DB	907	AAGACCCAGTCTTAGGGCCCCCGAGTGGAGCCCAAAATGTGCCCTTCGCACTTCTTC	966
QY	201	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg	220
DB	967	TTTCTCTTCATCAATTAACACGACGACTGTGGGTGCCACATCGAGACAGGAGGAGCGG	1026
QY	221	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly	240
DB	1027	CGGCTGACCTTTCGCCACCAAGGTTTATCCAAATGCTCGATGAGACCCCAAGTCTGGCGT	1086
QY	241	ValAlaThrPheValIleGlnGlnGluIlePheAspArgPheThrGlyTyrTrpTyrPro	260
DB	1087	GTGGCCACCTTCGTCATACACGGAAGAGTTTCGACCGCTTACTGGGTACTGGGTGGTGGCC	1146
QY	261	ThrAlaSerTrpGluGlySerGluGlyLeuLysTrpLeuArgIleLeuTyTrpGluVal	280
DB	1147	ACACCTTCCTGGGAAGGTTAGAGGGCTTCAGAGCCTCGAATCTCTGATATAGGAATTC	1206

QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
 Db 1207 GATGAGTCGCGAGGTGAGAGTACGTACGTCCTCTCGGCTTGAGAAAGAGAGACG 1266
 QY 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
 Db 1267 GACTCGTATCGGTACCCAGGACAGGACAAAGATCCCAAGATGCTGTAACCTGGCT 1326
 QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
 Db 1327 GAGTTCACAGCTGACAGCGACGAGGCAAGATCGTCCAGCCAGGACGAGAGAGCTGGTGAG 1386
 QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPheArgAsp 360
 Db 1387 CCGTTCACAGCTGCTTCCCGAGGTGAGTACATCCCGAGGCGCGGTGGACCGGGGAT 1446
 QY 361 GlyLysTyrAlaTyrPalaMetPheLeuAspArgProGlnGlnTyrPheGlnLeuValLeu 380
 Db 1447 GGCAGAAATACGCGCTGGGCGCATGTTCTGGACCGGCGCCAGAGTGGCTCCAGCTCGCTGC 1506
 QY 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAla 400
 Db 1507 CTCCCGCGCGCTGCTTCATCCCGAGCACAGAAATGAGGAGCAGCGCTACCTCTGCGC 1566
 QY 401 ArgAlaValProArgAsnValGlnProTyrValAlaTyrGluGluValThrAsnValTyr 420
 Db 1567 AGAGCTGTCCCGAGAGATGCCGCTATGTGGTGTACGAGGAGGTCCACACGCTGTGG 1626
 QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluLysAspLysLeuLys 440
 Db 1627 ATCATATGTTACATGACATCTTATCCCTCCCAATACAGAGAGAGAGAGAGAGCTGCGC 1686
 QY 441 PheLeuAlaGluAlaSerGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 460
 Db 1687 TTTCTCCCGCGCAATGCAATGCAAGACCGGCTCTCCATTTGTACAAAGTCCACCGCGTT 1746
 QY 461 LeuLysSerGlnGlyTyrAspTyrPheSerGluProPheSerProGluLysAspLysPheLys 480
 Db 1747 TTAATATCCCGAGGCTGACATTTGAGTGTAGCCCTTCAGCCCGGGGAGAGATATAATTTAAG 1806
 QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyGluTyrPglValIleuAlaArgHis 500
 Db 1807 TGCCCATTAAGAGAGAGAGATTGCTGTACACACGCGGTGATGGAGGTTTGGCGAGGCGC 1866
 QY 501 GlySerLysIleTyrValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
 Db 1867 GGCTCCAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTCCAGGCGCACCAAGGAC 1926
 QY 521 ThrProLeuGluHisIleLeuTyrValIleSerTyrGluAlaAlaGlyIleValArg 540
 Db 1927 AGCGCGCTGGAGCACACCTTACGTGTGCTACGTATGAGGCGCGCGGAGATCTGTACC 1986
 QY 541 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
 Db 1987 CTCACACGCGCGGCTTCTCCCATAGCTGCTCCATGAGCAGAACTTGACATGTGTTGCTGC 2046
 QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
 Db 2047 AGCCACTACAGCAGGTGAGCAGCGCGCGCTGTCAGCTGTACAAAGCTGAGCGCGCCC 2106
 QY 581 AspAspAspProLeuHisLysGlnProArgPheTyrPalaSerMetMetCysIleAlaIleSer 600
 Db 2107 GACGAGCAGACCCCTGCACAGACGCCCTTCTGGGCTAGCATGTGAGGACGACGAC 2166
 QY 601 CysProProAspTyrValProProGluLysPheHisPheHisThrArgSerAspValArg 620
 Db 2167 TGCCCCCGGAGTATGTCTCTCCAGAGATCTTCATTTCCACAGCGGCTCGGATGTGGCG 2226
 QY 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 640
 Db 2227 CTCTACGGCATGATCTACAAAGCCCAAGCGCTTGACGCGGAGGAGAAAGCACCACCGCTC 2286

QY 641 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
 Db 2287 CTCTTTTATATGAGAGGCCCCAGGTGACAGCTGGATTAATCTCTTCAAGAGCATACAG 2346
 QY 661 TyrLeuAlaGluAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArg 680
 Db 2347 TACTTGGCGGCTCAACACACTGCGCTCCCTGGGCTACGCGCGTGTGATTCAGCGGACG 2406
 QY 681 GlySerCysGlnArgGlyLeuArgPheGluGluAlaLeuLysAsnGluMetGlyLysVal 700
 Db 2407 GGCTCTGTACAGAGGCTTGTGTTCCAAAGGGGCTTGAAAAACCAATGGGCGGAGTG 2466
 QY 701 GluIleGluAspGluValGluGluLysLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
 Db 2467 GAGATCGAGAGACAGGTGAGGGGCTCGAGTTCTGTGGCGGAGAAATATGGCTTCATCCAG 2526
 QY 721 LeuSerArgValAlaIleHisGlyTyrPheTyrGlyGlyPheLeuSerLeuMetGlyLeu 740
 Db 2527 CTGAGCCGAGTGGCCATCCATGCTGTGTCTGACGGGGGCTTCTCTCGCTCATGGGCGTA 2586
 QY 741 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTyrPhe 760
 Db 2587 ATCCACAAAGCCCGAGGTGTCAAGG----- 2611
 QY 760 ValTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlu 780
 Db 2611 ----- 2611
 QY 780 YTrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLe 800
 Db 2612 -----CCCAACCGGCTT 2622
 QY 800 uLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAspPheLeuVal 820
 Db 2623 GCTTATCCCTCCAGCGGCTTCTGGAGGAAACGTGCACATTTTCCACAAACTTCTCGT 2682
 QY 820 LserGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 840
 Db 2683 CTCCCACTGATCCGAGCAGGAGGAAACCTTACCAAGCTCCAGATCTACCCCAAGAGACA 2742
 QY 840 sSerIleArgCysProGlnSerGlyLysHisTyrGluValThrLeuHisPheLeuGlu 860
 Db 2743 CAGTATTCGCTGCGCGCGAGATCGGGGAGCAGCATATGAAATGACAGTGTGCTTCTACA 2802
 QY 860 nGluTyrLeu 863
 Db 2803 GGAATTAACCTC 2812
 RESULT 8
 US-09-976-674-40
 ; Sequence 40, Application US/09976674
 ; Patent No. US20020115843A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Qi, Steve
 ; APPLICANT: Akinsanya, Karen
 ; APPLICANT: Riviere, Pierre
 ; APPLICANT: Julien, Jean-Louis
 ; TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPLY
 ; FILE REFERENCE: 70669
 ; CURRENT APPLICATION NUMBER: US/09/976,674
 ; PRIOR APPLICATION NUMBER: 2001-10-12
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 40
 ; LENGTH: 4037
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-976-674-40
 Alignment Scores: 0 Length: 4037
 Pred. No.:


```

QY 381 LeuProAlaLeuPheIleProSerThrGluAsnGluGluAlaArgLeuAlaSerAla 400
    |||||
Db 1507 CTCGCCCGCGCCGTGTCATCCCGACAGAGATGAGAGACGCGCTGCTCTGCC 1566

QY 401 ArgAlaValProArgAsnValGlnProIleValValTyrGluGluValThrAsnValTrp 420
    |||||
Db 1567 AGAGCTGTCGCCAGAAATGTCACGCCGTATGTGTACGAGAGAGGTUACCAACCTCTGG 1626

QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGluLeuGlyCys 440
    |||||
Db 1627 ATCATATGTTCAATGACATCTTCTATCCCTTCCCAATAGAGGAGAGAGAGCTCTGCC 1686

QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrValThrAlaVal 460
    |||||
Db 1687 TTTCTCCCGCCCAATGAAATGCACAGCCGCTCTGCCATTGTACAAAGTCACCCGCTT 1746

QY 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluGluAspGluPheLys 480
    |||||
Db 1747 TTTAAATCCAGGGCTACGATGGATGAGTGAGCCCTTCACGCCCGGGGAGATGAATTTAG 1806

QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyLysTrpGluValLeuAlaArgHis 500
    |||||
Db 1807 TGCCCATTAAGGAAGAGATTGCTGTACACAGCGGTGAATGGAGTTTGGCGAGGCAC 1866

QY 501 GlySerLysIleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAsp 520
    |||||
Db 1867 GGCTCCAG-----GGCACCAAGGAC 1887

QY 521 ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaAlaGluIleValArg 540
    |||||
Db 1888 AGCGCGGTGGAGCACACCTTACGTGGTACGTATGAGGGCGCGGAGATGCTGTACG 1947

QY 541 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 560
    |||||
Db 1948 CTCACACAGCCCGGCTTCCCATAGCTGCTCCATGAGCCGAACCTTGCATGTCTGC 2007

QY 561 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 580
    |||||
Db 2008 AGCCACTACACACACGTCGACACGCCCGCGCTGCACGCTCAAGCTGAGCGGCCCC 2067

QY 581 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer 600
    |||||
Db 2068 GACACAGACCCCTGCACACACACCCCGCTTGGGTACCATGATGAGGACGCCACG 2127

QY 601 CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 620
    |||||
Db 2128 TGCCCCCGGATATGTTCCCTCCAGAGATCTTCATTTCCACAGCGCTGCGATGTGCG 2187

QY 621 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysHisProThrVal 640
    |||||
Db 2188 CTCTACGCGCATGATCTACACACCCCGCTTGCACGCGAGGAGAAAGCACCCACGCTC 2247

QY 641 LeuPheValTyrGluGlyProGluValGlnLeuValAsnAsnSerPheLysGlyIleLys 660
    |||||
Db 2248 CTCTTTGTATATGAGGCGCCAGAGGTGACGTGTGAATTACTCTCTCAAGGCAATCAAG 2307

QY 661 TyrLeuArgLeuAsnTrpLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArg 680
    |||||
Db 2308 TACTTGGGGGTCACACACTGTGCTCCCTGGGCTACGCCCGGTGTGATGTAGAGGCGAG 2367

QY 681 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnVal 700
    |||||
Db 2368 GGCTCTGTCTAGCGAGGCTTCGTTGGAAGGGGCCCTGAATAAACCAATGGCCAGGTG 2427

QY 701 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 720
    |||||
Db 2428 GAGATCGAGGACAGGGAGGAGGCTGAGTCTGTGCGCGGAAGTATGGCTTATGTGAC 2487

QY 721 LeuSerArgValAlaIleHisGlyTyrSerTyrGlyLysPheLeuSerLeuMetGlyLeu 740
    |||||
Db 2488 CTGAGCGGAGTGTGCGCATGAGTGGTGTCCAGCGGGGCTTCTCTGCTCATGTGGGCTA 2547

```

```

QY 741 IleHisLysProGlnValPhe-LysValAlaIleAlaGlyAlaProValThrValTrpMe 760
    |||||
Db 2548 ATCCAAAGCCCGCAGGTGTCAAG----- 2572

QY 760 TalatyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGly 780
    |||||
Db 2572 ----- 2572

QY 780 TyrGluValAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLe 800
    |||||
Db 2573 -----CCCAACCGCTT 2583

QY 800 uLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuVal 820
    |||||
Db 2584 GCTTATCTCCACGCGCTCTCGAGCAAAAGCTGACACTTTTCCACACAAACTCTCTCGT 2643

QY 820 LserGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluValArgHis 840
    |||||
Db 2644 CTCCTCACTGATCCGAGCGAGGAAACCTTACAGCTCCAGATCTACCCCAACGAGAGACA 2703

QY 840 sSerIleArgCysProGlnSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGly 860
    |||||
Db 2704 CAGATTTCGCTGCGCCCGAGTGGCGGAGCAGCACTATGAAGTCACTGTGCTGCACTTTCTACA 2763

QY 860 nGluTyrLeu 863
    |||||
Db 2764 GGAATACCTC 2773

RESULT 10
US-09-976-674-2
: Sequence 2, Application US/09976674
: Patent No. US20020115843A1
: GENERAL INFORMATION:
: APPLICANT: Qi, Steve
: APPLICANT: Akinsanya, Karen
: APPLICANT: Riviere, Pierre
: APPLICANT: Junien, Jean-Louis
: TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPMV
: FILE REFERENCE: 70669
: CURRENT APPLICATION NUMBER: US 60/240,117
: PRIOR APPLICATION NUMBER: US 60/240,117
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 61
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 2671
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-976-674-2

Alignment Scores:
Pred. No.: 8, 28e-300
Score: 2870.00
Percent Similarity: 77.50%
Best Local Similarity: 61.55%
Query Match: 61.77%
DB: 10 Gaps: 2

US-09-976-674-3 (1-863) x US-09-976-674-2 (1-2671)

QY 24 PheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGlySerArgLys 43
    |||||
Db 110 TTTTATGTTGAGCGGATTTCTCTGAGTCACTTAATAAAGCTGTGCGGATACAGAAAA 169

QY 44 TyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysThrAsp 63
    |||||
Db 170 TATCATGGCTACATAGAGGCTTCCGACACACATCATTTTCATGTGTGGAAGAGAAATGAT 229

QY 64 GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlnLysMetProTyrGlySerArgGlu 83
    |||||
Db 230 CAGATGAGACTCATTTGACACAGAAATTAATTACCTTCCATGTGTGAGGAAACAGAAA 289

```


OY	84	AsnSerLeuLeuTyrSerGluIleProLysValArgLysGluAlaLeuLeuLeu	103
Db	290	AAATCACTGTTTAACTCGAAATTTCCCAAACTCAATAGAGCGAGCTTTAATGCTC	349
OY	104	SerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisGlyValTyrSerArg	123
Db	350	TCCTGGAAAGCCCTTTTGGATCTTTTTCAGGCAACCTGACTATGGAATGATTTCTGA	409
OY	124	GluGluIleLeuLeuArgGluArgLysArgLysGluValPheGlyIleThrSerTyrAsp	143
Db	410	GAAGAAAGACTATTTAAGAGAAAGAAAACCCATTGGACAGCTGGCAATTCCTTACAT	469
OY	144	PheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArg	163
Db	470	TATCAACCAAGAAAGTGGAAACATTTTCTGTTTCAAGCGGATGGAATTAATCACTGAAA	529
OY	164	AspSerGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIleLysThrGln	183
Db	530	GATGAGAGGGCCACAGAGATTTTACGCAACAACCTTTTAAGCCCAATCTAGTGGAAACTAGT	589
OY	184	CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe	203
Db	590	TGTGCCCAACATACGAGATGAGATCCAAAATTATGCGCTGCATCCAGACTGGATTCCTTT	649
OY	204	IleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArgLeuThr	223
Db	650	ATACATGACAACGATTTATTTGGATATCTAATCATCTGTAACAGAGAAAGAAAGCACTGCT	709
OY	224	PheCysHisGlnGlyLeuSerAsnValIleAspAspProLysSerAlaGlyAlaIleThr	243
Db	710	TATGTGCAACATGACTGACGCAACATGGAAGAAAGATCCAGATCAGCTGAGTGGCTACCT	769
OY	244	PheValIleGlnGluGluPheAspArgPhePheThrGlyTyrTrpTrpCysProThrAlaSer	263
Db	770	TTTGTTCTCCAAAGAAATTTGATTAATATCTCGCTATTTGGTGTGTCCTCAAAAGCTGAA	829
OY	264	TrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluValAspLeuSer	283
Db	830	ACAACCTCCCACTGGTGGT---AAAAATCTTAGATCTTATGAAAGAAATGATGAATCT	886
OY	284	GluValGluValIleHisValProSerProAlaLeuGluGluArgLysThrAspSerTyr	303
Db	887	GAGGTGGAATTTATCTCATGTTACATCCCTATGTTGGAAACAAAGAGGCGAGATTCAATTC	946
OY	304	ArgTyrProArgTrpArgSerLysSerAsnProLysIleAlaLeuLysLeuAlaGluPheGln	323
Db	947	CGTATCTCTAAACAGSTACAGCAAACTCTAAAGTACTTTTAAGATGTCCAGAAATATG	1006
OY	324	ThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGlnProPheSer	343
Db	1007	ATTGATGCTCGAAGAGATCATATGATGTGCATGATGAATGAAGCACTAATCAACCTTTGAG	1066
OY	344	SerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyr	363
Db	1067	ATTCTATTGGAAGAGATTGAATATATGTCACAGCTGATGAGACTCTGAGAGGAAAATAT	1126
OY	364	AlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProPro	383
Db	1127	GCTTGTCGCAACCTCAATAGTCCGCTCCCACTCCGCGCTCAATATGTTGATCTCACTT	1186
OY	384	AlaLeuPheIleProSerThrGluAsnGluGluArgLeuAlaSerAlaArgAlaVal	403
Db	1187	GAATTAATTTATCCAGTAAGAAATGATGTTATGAAAGACGACACTAATGAGTCACTG	1246
OY	404	ProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrpIleAsnVal	423
Db	1247	CCTAATTTCTGAGCGCACTAAATTAATCTATGAAGAAACAAACAGCATCTGATTAATATTC	1306
OY	424	HisAspIlePheTyrProPheProGlnSerGluGlyGluAspGlnLeuCysPheLeuArg	443
Db	1307	CATACATCTTTTCAATGTTTTTCCCAAAAGTCAC---GAAGAGGAATTAATGATTTATTTT	1363
OY	444	AlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSer	463

Db	1364	GCCTCTGAATGCAAAACAGGCTTCCGTCATTTATACAAATTAACATCTAATTTAAAGGA	1423
Oy	464	GlnGlyIyrAspTrpSerGluProPheSerProGlyIgluAspGluPheLysCysProIle	483
Db	1424	AGCAAAATTAACGATCCAGTGGTGGGCTCCGCTCCAAAGTGATTTCAAGTGCTATTC	1483
Oy	484	LysGluGluIleAlaLeuThrSerGlyIgluTrpGluValIleAlaArgHisGlySerLys	503
Db	1484	AAACAGAGATAGCAATTTACCAGTGGTGAATGGGAAGTCTTGCGCGGATGGATCTAAT	1543
Oy	504	IleTrpValAsnGluIuThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeu	523
Db	1544	ATCCAGATGTCAGAGTCACAAAGAGCTGGTATATTTTGAAGGACCAACAAAGATCCCTTTA	1603
Oy	524	GluHisIleLeuValValSerTyrGluAlaIleGlyIleValArgLeuThrThr	543
Db	1604	GAGATACCTCGTACGTAAGTCAGTTACGTAAATCCGGAAGGTGACAAAGGCTGACTAC	1663
Oy	544	ProGlyPheSerHisSerCysSerMetSerGluAsnPheAspMetPheValSerHisTyr	563
Db	1664	CGTGGCTACTACACTTTCTTGCTGCATGACGACACCTGCACTCTTTTAAGTAAGTAT	1723
Oy	564	SerSerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAsp	583
Db	1724	AGTAACCAAGAAATCCACACACTGNGTCCCTTTACACGTAATCAAGTCTCAAGATGAC	1783
Oy	584	ProLeuHisLysGlnProAlaGhrPheThrPheSerMetMetGluAlaIleAspCysPro	603
Db	1784	CCAACTTGCAAAACAAAGAAATTTTGGGCCACATTTTGAAATTCAGCAGCTCTTCTCT	1843
Oy	604	AspTyrValProProGluIlePheHisPheHisThrArgSerAspValArgLeuTyrGly	623
Db	1844	GACTATACCTCTCCAGAAATTTTCTCTTTGAAAGTACTACTGGATTACATTGATATGG	1903
Oy	624	MetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheVal	643
Db	1904	ATGCTCTACAAAGCCTCATGATCTACAGCCTGGAAGAAATATCCATCTGCTGTTCTATA	1963
Oy	644	TyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArg	663
Db	1964	TATGTGTGCTCTCAGGCGCATGGTGCCTAAATTCGATTTAAAGAGTCAAGATTTTCCGC	2023
Oy	664	LeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlyLysCys	683
Db	2024	TTGAATACCTCAGCCTCTCTTAGGTATAGTGTGTGTATGATACACACAGGGGATCTCTGT	2083
Oy	684	GlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGlu	703
Db	2084	CACCGAGGCGCTTAATTTGAAGGCGGCTTTAAATTAATTAATGGGTCGCAATACAAATTGAC	2143
Oy	704	AspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArg	723
Db	2144	GATCAGGTGGAGAGCACTCAATATCTAGCTTCGTGATATGATTAATTCATGACTTAATAGTGT	2203
Oy	724	ValAlaIleHisGlyTyrPserTyrGlyIgluPheLeuSerIleMetGlyLeuIleHisLys	743
Db	2204	GTTGGGCACTCCAGGCTGTCTCTTAGGATGAGATACCTCTCCCTGATGGCATTAATGCAAGG	2263
Oy	744	ProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAsp	763
Db	2264	TCAGATATCTTCAGAGTGTGCTATGTGGGGGCCCAAGTCAGTCTGGATCTTCTATGAT	2323
Oy	764	ThrGlyTyrThrGluAlaGlyTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAla	783
Db	2324	ACAGGATACACGGAACCTTATATGGGTCAACCTTCGACCAAGAACAGGAGGCTATTACTTA	2383
Oy	784	GlySerValAlaLeuHisValGlyLysLeuProAsnGluProAsnArgGluLeuIleLeu	803
Db	2384	GGATCTGTGGCATGCAAGCAAGAAAGTTCCTCTCGAAACCAAAATGTTTACTGCTCTTA	2443
Oy	804	HisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeu	823

QY	544	ProGlypheSerHisSerCysSerMetSerGlnAsnPhneAspMetPheValSerHisTyr	563
Db	1870	CGTGGTACACATCTTCCTGCTGCACAGCTGAGCTGGACTTCCTTTTATTAAGACATAT	1929
QY	564	SerSerValSerThrProProCysValHisValTyrTyrLysLeuSerGlyProAspAsp	583
Db	1930	AGTAAACAGAAAGATCCACACTGTGTGCTCCTTTACAACTATACAGTCCGGAAGATGAC	1989
QY	584	ProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaLeuSerCysPro	603
Db	1990	CCAACCTGCAAAACAAAGAAATTTTGGCCACACATTTTGATGTACGAGGTCTTCCT	2049
QY	604	AspTyrValProProGluLeuPheHisPheHisStnArgSerAspValArgLeuTyrGly	623
Db	2050	GACTATACTCTCCAGAAATTTTCTTTTGAAGACTACGATTTACATTTGATGGG	2109
QY	624	MetLeuTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheVal	643
Db	2110	ATGCTCTACAAAGCCTATATATCTACAGCCTGGAAAGAAATATCCTACTGTGCTGTTCATA	2169
QY	644	TyrGlyGly--ProGlnValGlnLeuValAsnAsnSerPheLysGlyLeuSerTyrLeuA	663
Db	2170	TATGTGTGTCTCTCAGGTGCTCAGTGTGATTAATTCGTTTAAAGAGTCAAGATTTTCC	2229
QY	663	rgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValAlaLeuAspLysArgLysSer	683
Db	2230	GCTTGAATACCCCTAGCCTCTTAGTATATGTGTGTATGATGTAAACAAAGGGGATCTT	2289
QY	683	ysGlnArgLysLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGluLeuG	703
Db	2290	GTCACGAGGGCTTAATTTGAAGGGGCTTTAAATATTAATTTGGTCAATATAGAAATG	2349
QY	703	LysArgLysValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheLeuAspLeuSerA	723
Db	2350	ACGATACAGGTGAGAAAGCATCTCAATATCTAGCTTCGATATGATTTCTACTACATTAGATC	2409
QY	723	rgValAlaLeuHisGlyTyrSerTyrGlyGlyPheLeuSerLeuMetGlyLeuLeuHisL	743
Db	2410	GTGTGGGCATATCCAGGCTGGTCTCTATGAGAGATACCTCTCCCTGATGGCATTTAATGCAG	2469
QY	743	ysProGlnValPheLysValAlaLeuAlaGlyAlaProValThrValTrpMetAlaTyrA	763
Db	2470	GGTCAGATATCTTACGGGTTCGATATGGTCGTGGGCCCAATCACTGTGGATCTTTATG	2529
QY	763	spThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluA	783
Db	2530	ATACAGATATACAGGAACGTTATATGGTCACCCCTGACAGATGAACAGGGCTATATCT	2589
QY	783	lagLysSerValAlaLeuHisValGlnLysLeuProAsnGlyProAsnArgLeuLeuLeu	803
Db	2590	TAGATCTGTGGCCATCAACAGCAAGAAAGTCCCTCTGAAACCAATGCTTATCTGCTCT	2649
QY	803	eunHisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnL	823
Db	2650	TACATGGTTCCTGGATGAGAAATGTCATTTGGACATATCCAGTATATTACTAGAGTTT	2709
QY	823	eunLeuArgAlaGlyLysProTyrGlnLeuGlnLLeuTyrProAsnGluArgHisSerLeuA	843
Db	2710	TACGAGAGGGCTGCAAAAGCCATATGATTTACAGATCTATCCTGAGAGACACAGCAATA	2769
QY	843	rgCysProGlnSerGlyGlnHisTyrGlnValThrLeuLeuHisPheLeuGlnGlyLysL	863
Db	2770	GAGTTCCTGATCGGAGAACATTTATGAACCTGCATCTTTTGACATCTTCAAGAAAC	2829
QY	863	eu 863	
Db	2830	TT 2831	
RESULT 12			
US-09-976-674-22			
; Sequence 22, Application US/09976674			
; Patent No. US20020115843A1			

Alignment Scores:			
Pred. No.:	4,166-271	Length:	4676
Score:	2608.00	Matches:	476
Percent Similarity:	72.26%	Conservative:	131
Best Local Similarity:	56.67%	Mismatches:	181
Query Match:	56.13%	Indels:	53
DB:	10	Gaps:	3
US-09-976-674-3 (1-863) x US-09-976-674-20 (1-4676)			
QY	24	PheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGlySerArgLys	43
DB	316	TTTTATGTTGAGCGGATTCCTGGAGTCAGCTTAAAGCTGCTGCCATACCAAGAAA	375
QY	44	TyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysTrpAsp	63
DB	376	TATCATGGCTACATGTGGCTAAGCCACACATGATTCATGTGTTGTAAGAGGAATGAT	435
QY	64	GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyLysProTyrGlySerArgLys	83
DB	436	CCAGATGGACCTCATTCACACAGAAATCTATTACCTTGCCATGCTGCTGGAGACAGAGA	495
QY	84	AsnSerLeuLeuTyrSerGlyIleProLysValArgLysGlnAlaLeuLeuLeuLeu	103
DB	496	AATACACTGTTTATCTGAAATTCACAAACATCATAGACACAGACTTAATGCTC	555
QY	104	SerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyValTyrSerArg	123
DB	556	TCTTGGAAACCTCTTTTGGATCTTTTCAAGGCAACACTGACATNTGAAATGTATCTCGA	615
QY	124	GlnGlnGlnLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThrSerTyrAsp	143
DB	616	GAGAGAACTATTATAGAGAAAGAAACCATTTGAGACAGTCGGAATGCTTCTTACGAT	675
QY	144	PheHisSerGlySerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArg	163
DB	676	TATCACCAAGAAAGTGAACATTTCTGTTTCAAGCCGGTAGTGAAATTATATACGTAANA	735
QY	164	AspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIleLysThrGln	183
DB	736	GATGGAGGCCACACAGAAATTTACGCAACAACCTTTAAGGCCCAATCTAGTGAACATAGT	795
QY	184	CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe	203
DB	796	TGTCCCAACATACGATGATGATCCAAAATATATGCCCTGCTGATCACAACGTGGATTGCTTT	855
QY	204	IleAsnAsnSerAspLeuTyrValAlaAsnIleGluThrGlyGlnIleuArgArgLeuThr	223
DB	856	ATACATAGCAGACGATATTGGATATCTAACATCGTAACAGAGAAAGAGACATCAGT	915
QY	224	PheCysHisGlnGlnLysLeuSerAsnValIleuAspAspProLysSerAlaGlyValAlaThr	243
DB	916	TATGTCCACAAATGAGCTAGACATGAGGAAGATGCCAGATCAGCTGAGTGGCTACC	975
QY	244	PheValIleGlnGlnLysLeuPheAspArgPheThrGlyTyrTrpTyrCysProThrAlaSer	263
DB	976	TTTGTTCTCCAGAAATAATTTGATAGATATTCGGCTATTGGTGGTGTCCAAAAGCTGAA	1035
QY	264	TrpGlnGlySerGlnLysLeuLysTrpLeuArgIleLeuTyrGlnIleuValAspGluSer	283
DB	1036	ACAATCCCAAGTGGTGGT---AAATTCCTAGAAATTCCTATATGAGAAATGATGATCT	1092
QY	284	GluValGlnValIleHisValProSerProAlaLeuGlnIleuArgLysTrpAspSerTyr	303
DB	1093	GAGGTGGAATTTATTCATGTTTACATCCCTATCTGGAAACAGGAGGCGAGATTATCT	1152
QY	304	ArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGlnPheGln	323
DB	1153	CGTTATCTTAAACAGAGTACAGCAAAATCTTAAAGTCACTTTTAAAGATGTAGAAATATAG	1212
QY	324	ThrAspSerGlnLysIleValSerThrGlnGlnLysGlnLeuValGlnProPheSer	343

DB	1213	ATTGATGCTGAAGGAAGATCATAGATGTCAATAGATAAGAACTAATTCACCTTTTGAG	1272
QY	344	SerLeuPheProLysValIleuTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyr	363
DB	1273	ATTCTATTGGAAGAGATTGAATATATTTGCCAGAGCTGGATGGATCTCTGAGGGAATAAT	1332
QY	364	AlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnIleuValLeuLeuProPro	383
DB	1333	GCTTGTCATCCTACTATAGATGCTCCAGACGCGCTCTACAGATGATGTTGATCTACCT	1392
QY	384	AlaLeuPheIleProSerThrGlnAsnGlnGlnIleArgLeuAlaSerAlaArgAlaVal	403
DB	1393	GAAATTATTATCCCAAGATGATGTTATGAGAAAGCAGACAGACTCATTTGAGTCACTG	1452
QY	404	ProArgAsnValGlnProTyrValValTyrGlnGlnIleuThrAsnValTrpIleAsnVal	423
DB	1453	CGTATTCGTGCGCCACTAATTAATTCATGAGAAACAACACATCTGGATAAATATC	1512
QY	424	HisAspIlePheTyrProPheProGlnSerGlnGlyLysAspGlyLeuLysPheLeuArg	443
DB	1513	CATGACATCTTTCATGTTTCCCAAGTCAAC---GAAAGAGAAATTCAGTTATTTT	1569
QY	444	AlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSer	463
DB	1570	GCCTGGAATGCAAAACAGGTTTCCGTCATTTATACAAATATTCATCTAATTTTAAAGAA	1629
QY	464	GlnGlyTyrAspTrpSerGlnProPheSerProGlyGlnAspGlyLeuLysCysProIle	483
DB	1630	AGCAAAATATAAAGCATCCAGTGTGGCTGCCCTCCCAAGTGAATTCAGATGTCCTATC	1689
QY	484	LysGlnGlnIleAlaLeuThrSerGlyGlyTrpGlnIleuValLeuAlaArgHisGlySerLys	503
DB	1690	AAAGAGAGATGACATTTACACAGTGTGAATGGAAATCTTGTGGCGCATGGATGTAAT	1749
QY	504	IleTrpValAsnGlnIleuThrLysLeuValTyrPheGlnIleuThrLysAspThrProLeu	523
DB	1750	ATCCAGATGATGATGACAGAAAGCTGGTATATTTTGAAGGCCAACAAGCTCCCTTAA	1809
QY	524	GlnHisHisLeuTyrValValSerTyrGlnAlaIleGlyGlnIleuValArgLeuThrTrp	543
DB	1810	GAGCATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1869
QY	544	ProGlnPheSerHisSerCysSerMetSerGlnAsnPheAspPheValSerHisTyr	563
DB	1870	CGTGGTACGACATCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT	1929
QY	564	SerSerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspAsp	583
DB	1930	AGTAACACAGAAAGATCCACACTGTGTGCTCTTACAAAGCTATCAAGTCTGTAAGATGAC	1989
QY	584	ProLeuHisLysGlnProArgPheThrAlaSerMetMetGlnAlaIleSerCysProPro	603
DB	1990	CCAACTTGCAGAAACAAAGGAATTTTGGCCACCAATTTTGGATTCACAGGCTCTTCC	2049
QY	604	AspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArgLeuTyrGly	623
DB	2050	GATATATCTCTCCAGAAATTTCTCTTTGAAGAAGTACTGATTTACATTTGATGATG	2109
QY	624	MetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValIleuPheVal	643
DB	2110	ATGCTCTACAAAGCTCATGATCTACAGCCCTGGAAGAAATATCTCATGCTGCTCAT	2169
QY	644	TyrGlnGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleuLysTyrLeuArg	663
DB	2170	TATGTGTGCTG-----	2180
QY	664	LeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCys	683
DB	2180	-----	2180
QY	684	GlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGlnIleGln	703
DB	2181	-----GTCGCAAAATAGAAATTTGAC	2198

QY 704 AspGlnValGluGlyLeuGlnPheValAlaGluTyrGlyPheIleAspLeuSerArg 723
|||||.....
Db 2199 GATCAGGGGAAAGACTCCAAATATCATGCTTCATATGATTTCATGACTAGATGCT 2258
QY 724 ValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLys 743
|||||.....
Db 2259 GTGGGCAATCCACGGGCTGGCTTATGAGAGATACCTCTCCATGAGCATTAATGCAGAGG 2318
QY 744 ProGluValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetLeuTyrAsp 763
|||||.....
Db 2319 TCAGATATCTTCAGGGGTGCTATGCTGGGCCCCAGTCCAGTCTGGATCTTCTATGAT 2378
QY 764 ThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAla 783
|||||.....
Db 2379 ACAGATACACGGAACGTTATATGTTGGTCCACCTGACAGAAATGACAGGGCATTA 2438
QY 784 GlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeuLeuIleLeu 803
|||||.....
Db 2439 GGATCTGGGCGCATGCAGAACGAAAGTTCCTCTGAAACCAATCGTTTACTGCTCTTA 2498
QY 804 HisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeu 823
|||||.....
Db 2499 CATGCTTCCTGGATGAGATGTCATTTTGGACATACCATGATATTTACTGAGTTT 2558
QY 824 IleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHisSerIleArg 843
|||||.....
Db 2559 GTGAGGGCTGGAAAGCCATATGATTACAGATCTATCTCAGAGAGACACAGCATAGA 2618
QY 844 CysProGluSerGlyGlnHisTyrGluValThrLeuLeuHisPheLeuGlnIleTyrLeu 863
|||||.....
Db 2619 GTTCTGTAATGGGAGAACATTTATGACTGATCTTTTGGACTACTTCAGAAACCTT 2678

RESULT 14
US-09-976-674-26
Sequence 26, Application US/09976674
Patent No. US20020115843A1
GENERAL INFORMATION:
APPLICANT: Qi, Steve
APPLICANT: Akinsanya, Karen
APPLICANT: Riviere, Pierre
APPLICANT: Junien, Jean-Louis
TITLE OF INVENTION: NOVEL SERINE PROTEASE GENES RELATED TO DPTIV
FILE REFERENCE: 70669
CURRENT APPLICATION NUMBER: US/09/976, 674
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240, 117
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Seq ID NOS: 3.1
SEQ ID NO 26
LENGTH: 2411
TYPE: DNA
ORGANISM: Homo sapiens
US-09-976-674-26

Alignment Scores:
Pred. No.: 4e-268 Length: 2411
Score: 2576.00 Matches: 503
Percent Similarity: 83.39% Conservative: 14
Best Local Similarity: 81.13% Mismatches: 54
Query Match: 55.45% Indels: 49
DB: 10 Gaps: 6

US-09-976-674-3 (1-863) x US-09-976-674-26 (1-2411)

QY 1 MetalThrThrGlyThrProThrAlaAspArgLysAspAlaAlaThrAspAspPro 20
|||||.....
Db 367 ATGGCACCACCGGAGCCCAACAGGCGACGAGCGACGACGCGCCACAGATACCCG 426
QY 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
|||||.....
Db 427 GCCGCCGCTTCAGAGTGCAGAAAGCATCTGGGACGGGCTCCGAGATCATCCACGGC 486

QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
|||||.....
Db 487 AGCCCAAGTACTCGGGCTCTATTTCACAAAGGCGCCCAAGACTTCAGATTGTGCAG 546
QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 80
|||||.....
Db 547 AAGAGGATGAGTGGGCGCCACCTCCACCGCTCTACTACTGGGAATGCATATGCG 606
QY 81 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysValArgLysGlnAlaLeu 100
|||||.....
Db 607 AGCCGAGAACTCCCTCTTACTCTGAGATTCCAAGAAAGTCCGGAAGAGGCTCTG 666
QY 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisGlyVal 120
|||||.....
Db 667 CTGCTCTGCTCTGAGACAGATGCTGATCTTTCCAGGCCACGCCACCATGGGGTC 726
QY 121 TyrSerArgGluGluGluLeuLeuArgLysArgLeuGlyValPheGlyIleThr 140
|||||.....
Db 727 TACTCTCGGAGAGAGAGCTGCTGAGGAGCGAAACCGCTGGGCTTCGGCATCAC 786
QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
|||||.....
Db 787 TCTACGACTTCCACAGGAGAGTGGCTTCTCTCTCCAGGCCACGACAGCTTTC 846
QY 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
|||||.....
Db 847 CACTCCCGGAGAGGGGAGAGAGCGCTTCATGATGCTCCCATATGAACCGCTGGAATC 906
QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
|||||.....
Db 907 AAGACCCAGTGTCTAGGGGCCCGGATGAGACCCCAAAATCTGGCCGACCTCTTC 966
QY 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 220
|||||.....
Db 967 TTCTCTTATCATTAACAGGACGCTGTGGGCAACATCGAAGACAGGAGGAGGG 1026
QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
|||||.....
Db 1027 CGGCTGACCTTCTGCGACCAAGGTTTATCCAAATGCTCGAATGACCCCAAGCTCGGGT 1086
QY 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCysPro 260
|||||.....
Db 1087 GTGGCCACCTTCTGTCATACAGGAAGAGTTCGACCCCTTCTACGTGGTGGTGGCC 1146
QY 261 ThrAlaSerTrpGluGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGluVal 280
|||||.....
Db 1147 ACAGCTCTCTGGAGAGTTCAAGGGCTCCAAAGACGCTGCGAATCTGTATGAGAAAGTC 1206
QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArgLysThr 300
|||||.....
Db 1207 GATGATCCGAGAGTGGAGTATTACGCTCCCTCTCTGCGCTGAGAAAGAAAGACG 1266
QY 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320
|||||.....
Db 1267 GACTCTGATCGGTACCCAGGACAGGCAAGAAATCCCAAGATTCCTTGAATGCGCT 1326
QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
|||||.....
Db 1327 GAGTTCAGACTGACAGGAGGCAAGATCGTCTGACCCAGGAAAGAGTGGTGGAG 1386
QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrTrpThrArgAsp 360
|||||.....
Db 1387 CCTTCACGCTGCTGTTCCTCCGAAGTGAATACATCGCCAGGCGCGGCTGACCGGGAT 1446
QY 361 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 380
|||||.....
Db 1447 GGCAAATACGCGCTGGGCATGTCTTGAGACGGGCCAGCAGTGGCTCCAGCTGCTCTC 1506
QY 381 LeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAla 400
|||||.....
Db 1507 CTCGCCCGGCGCTGTTCATCCCGAGCACAGAGATGAGAGACAGGCTGACCTTCGCC 1566


```

QY 401 ArgAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrp 420
Db 1567 AGAGTGTCCCGCAGGAAGTCCAGCCGTATGTGTGTAGAGAGAGTCCACCAAGCTGTGG 1626
QY 421 IleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGluLeuGly 440
Db 1627 ATCAATGTCATGACATCTCTTATCCCTCCCTCCCATCAGAGAGAGAGAGAGAGCTCTGC 1686
QY 441 PheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValAlaVal 460
Db 1687 TTTCTCCCGCCCAATGAATGAAGAACCGGCTTCTCCATTTGTACAAATCTACCGCCGTT 1746
QY 461 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGluGluAspLupheLys 480
Db 1747 TTAATAATCCAGAGGCTACGATTGGAGTAGACCTTCAGCCCGGGGGAAGTAGACAGAGC 1806
QY 481 -----CysPro 482
Db 1807 CTGACGAATGCTGTGACATCGCGGTAGTACGTGTGCTTCAATATGCTGTGTGTCA 1866
QY 483 Ile-----LysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArg 499
Db 1867 TTGGTGGCGCCCGCCACTCAGCCAGCACCCCTGGGGGAAGAAACAGAGGATCGGCAGG 1926
QY 500 HisGlySer-----LysIleTrpValAsnGluGlu-ThrLysLeuValTyr 514
Db 1927 AAGCAGAGCTTCCCGCAGTACTGCATGATGTGCGAGGGCTTAGAGCACCACCTGTGGC 1986
QY 514 rPheGlnGlyThrLysAspThrProLeuGluHisLeuTyrValValSerTyrGluAl 534
Db 1987 TTATTTCAGGACAGCATTTACT-----GAGCAGCCTCCCTGTGCCAGG----- 2029
QY 534 AlaLeuGluIleValAlaArgLeuThrThrProGlyPheSerHisSerCysSerMetSer 554
Db 2030 -----CCCTTAGCACACAGGAGGTTGGCCACCTACAGGCCACACAGTCA 2073
QY 554 AsnAspPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHisVa 574
Db 2074 A-----ATCCGCCACACCACTGTGTTCATTA 2100
QY 574 LTrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAla 593
Db 2101 ATAAAGTTTATTATGGCAGTGTAGCCACAGCCACTTGTTCACAGAGACTGTGTGGTGC 2158

RESULT 15
; Sequence 843, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 843
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 2643110CB1
US-10-044-090-843

Alignment Scores:
Pred. No.: 3,32e-265 Length: 2079
Score: 2548.00 Matches: 497
Percent Similarity: 84.67% Conservative: 11
Best Local Similarity: 82.83% Mismatches: 43
Query Match: 54.84% Indels: 49
DB: 12 Gaps: 6

```

```

US-09-976-674-3 (1-863) x US-10-044-090-843 (1-2079)
QY 1 MetaIaThrThyGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 20
Db 330 ATGGCCACCACCGGAGCCCAACGCGCAGATCGACGAGCGCGCCACACATACACCG 389
QY 21 AlaAlaArgPheGluValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
Db 390 GCCGCCCGCTTCAGTGCAGAGACACTGTGTGGAGCGGCTCCGAGCATCTCCACGCG 449
QY 41 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
Db 450 ACCCCCAAGTACTCGGGCTCATTTGCAACAGAGCGCCACAGCATCTCCAGTTTGGCAG 509
QY 61 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGluYmetProTyrGly 80
Db 510 AAGAGGAGTACTGGGGCCCACTCCACCCCTCTTACTGTGGAAATGGCATATGGC 569
QY 81 SerArgLysAsnSerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeu 100
Db 570 AGCCGAGAGACTCCCTCTTACTGTGATTTCCAAAGAGTCCGGAAGAGGCTGTG 629
QY 101 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120
Db 630 CTGCTCCTGTCTCGAAGCAGATGCTGATTCATTTCCAGGACAGCCACCATGGGCTG 689
QY 121 TyrSerArgGluGluGluLeuLeuArgGluValArgLysArgLeuGlyAlaPheGlyIleThr 140
Db 690 TACTCTCGGAGAGAGAGTGTGTGAGGAGGAGGAAAGGCTGGGGTCTTGGCAACAC 749
QY 141 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
Db 750 TCTTACGACTTCCACAGCAGAGAGTGGCTTCTTCTTCCAGGCCAGCAACGCTTTC 809
QY 161 HisCysArgAspGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 180
Db 810 CACTGGCGGAGGCGCGGCAAGAGCGCTTCATGTGTCCCTCATGAAACCGCTGGAATTC 869
QY 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
Db 870 AAGACCCAGTGTCTCAGGCGCCGAGATGAGCCCAAAATGTGCCCTGCCACCTGACTTC 929
QY 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrLysGluGluArg 220
Db 930 TTCTCCTTATCATATACACGACCTGTGGGTGGCCAAATGACAGACAGCGAGCGG 989
QY 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240
Db 990 CGGCTGACCTTGTGCCACCAAGGTTATGCCAATGTCTGGATGACCCCAAGTCTGGGGGT 1049
QY 241 ValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpTrpCysPro 260
Db 1050 GTGGCCACCTTGTCTATACAGGAAGAGTTGCAGCGCTTACTGTGGTGTGGTCCCC 1109
QY 261 ThrAlaSerTrpGluLysSerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluVal 280
Db 1110 ACAGCTCTCTGGAGAGTTTCAGAGGCTTCAGAGCCTCGAATCTGTATGAGAGATTC 1169
QY 281 AspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluValArgLysThr 300
Db 1170 GATGAGTCCGAGGTGAGGTCAATTCAGTCCCTCTCTGCGGCTAGAACGAAGAGAGC 1229
QY 301 AspSerTyrArgTyrProArgThrGlySerLysAspProLysIleAlaLeuLysLeuAla 320
Db 1230 GACTCTATCTGATCCAGGACGAGCAGCACAACAACTCCAAAGATTCCTTGAAGAGGCT 1289
QY 321 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGln 340
Db 1290 GAGTTCAACACTGACAGCAGGAGGATGCTGTCCAGCCAGGAGAGAGAGCTGTGGCAG 1349
QY 341 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgAsp 360
Db 1350 CCTTACGCTCGCTTCCGAGAGGTGAGTACATCGCCAGGCGCGGTGAGACCCGGAGT 1409

```


[illegible]

Db	620	ptHrProIleuIysValAlaIacIylProIeuthrArgValIleuAsnValAlaIaGluIleHisAs	640
QY	1756	CGAG-----CCCGTCGACAGAGAGCCCGCTTCTGGGCTAGCATGATGAGGAGC	1806
Db	640	pAspSerIleuGluInleuThrIysAspGluIysPheIysGluIysIle-----	656
QY	1807	CAGTCGCCCCCGATATATGCTTCTCCAGAGATCTTCATTCACACAGCC-----	1858
Db	657	-----LysAsnIyrAspLeuPro--IleThrSerIyrIysTrpMetValIleuAs	672
QY	1859	-TCGATGTGCGGCTGTACGCGCATGTATCTACAAGCCCCCGCTTCGACCGAGAGA	1917
Db	672	pAspGlyValGluIleAsnIyrIleGluIleIysProAlaAsnIleuAsnProIylsIly	692
QY	1918	GCACCCACACGCTCTCTTGTATATGAGAGGCCCGCCAGGTGCAGTGGATATACCTCT	1977
Db	692	stYrProIleuValAsnIleIyrGlyArgProGlySerGln-----	706
QY	1978	CAAGGCATCAAGTACTGGGGCTCAACACATCGGCTCCCGGGCTACGCGCTGTGT	2037
Db	707	-----ThrPheThrIyrSerSerIleuAlaPheGluGlnAlaIaVa	720
QY	2038	GATTGACGCG-----AGGGCTCCTGTGACGCGAG	2067
Db	720	IValSerGlyIleuAsnProAlaIleValIleuGlnIleGluProArgIylThrGlyIysG	740
QY	2068	GCTTCGGTTCGAGAGGGCCCTGAAAAACCAATGGCGCCAGGTGGAGATCGAGACCAGT	2127
Db	740	yTrpSerPheArgSerTrpAlaArgIuIysLeuGlyIyrTrpGluProArgAspIleTh	760
QY	2128	GGAG-----GGCGTCAGTTCGCGCGGAGAGATGTGCTTCACGACGTGAGCGGATGC	2184
Db	760	rGluValThrIlysIysPheIleGlnArgAsnSerGlnHisIleAspGluSerIysIleAl	780
QY	2185	CATCCATGGCGGTCTCAGAGGGGGCTTCTCTGCTCATGAGGGAGTA--ATCCACACACC	2241
Db	780	aIleIrrpGlyTrpSerIyrGlyGlyIlePheTrsIleuIysTrpValGluIleuAspAsnG	800
QY	2242	CGAGGTGTCAAGTGGCCATTCGCGGGTGGCCCGGACCGCTCGATGGCTCGACAC	2301
Db	800	yAspThrPheIysIyrAlaMetAlaValAlaIrrProValIrrAsnTrpIrrLeuIyrAspSe	820
QY	2302	AGGTCATCTAGCGGTACATGTGACGCTCCGTGAGAACACACGACGCGTGAAGCGGG	2361
Db	820	rValIyrThrIleuArgIyrMetAsnGlnProSerGluAsnHisGluGlyIyrPheGluIa	840
QY	2362	TTCCGTGGCCCTGCACGTGGAGAAAGTGGCCCAATGAGGCCAACCGCTGTACTCTCA	2421
Db	840	IserThrIleGlnAsnIrrPheIysSerPhe--GluSerLeuIysArgIlePheIleValI	859
QY	2422	CGGCTCTGTGACAAAAGCGACACTTTTCCACACAAACTCTGCTGTCCCAACGTAT	2481
Db	859	sgIlyThrPheAspAsnIrrHisIleGlnAsnTrpIrrPheArgIleValAspGlnIleuAs	879
QY	2482	CGAGACAGG---AAACCTTACAGACTCCAGATCTACCCCAACGAGACACAGTATTCG	2538
Db	879	nIleuIleuIyIleuThrAsnIyrAspMetHisIlePheProAspSerAspHisSerIleAr	899
QY	2539	C 2539	
Db	899	g 899	

Search completed: December 12, 2002, 11:53:34
Job time : 83.5 secs

J. Cell Biol. 108, 1363-1373, 1989
 A:Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an in
 A:Reference number: A30107; PMID:89174971; PMID:2647766
 A:Accession: A30107
 A:Molecule type: DNA
 A:Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRLET', 189-199, 'N', 201-365, 'DFRKRRKR', 376-57
 A:Cross-references: EMBL:X15484
 A:Note: the authors translated the codon ACC for residue 572 as Asn
 C:Genetics:
 A:Gene: SOD: DAP2; SRE13; MIPS:YHR028C
 A:Cross-references: SGD:S0001070; MIPS:YHR028C
 A:Map position: 8R
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
 F:30-45/Domain: transmembrane #status predicted <TMM>
 F:63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	4,24e-18	Length:	818
Score:	432.50	Matches:	164
Percent Similarity:	39.47%	Conservative:	91
Best Local Similarity:	25.39%	Mismatches:	262
Query Match:	9.02%	Indels:	129
DB:	1	Gaps:	28

US-09-976-674-4 (1-2617) x A30107 (1-818)

```

OY 728 GGGTGGGCCCTTCGTCATACAGAAAG---TTGACCGCTTCACGTGGTACTGGTG 784
DB 234 GLLYSPROASPRIPALTYRGLUGLIVLALPHEGLUASPLYSALAAATPTPT 253
OY 785 TGCCACACACCTCTCGGAAGGTTCAGAGGCGCTCAAGACGCTGCATCTGTATGAG 844
DB 254 SerProGlyAspTyr-----LeuAlaPheLeu 263
OY 845 GAAGTGATGATGCCGAGGTGAGGTCACTACGCTCCCTCTCTCGGCTAGAGAAAG 904
DB 264 LysIleAspIleSerGluValIleGluPheIleProTyrTyrValIleAspGluLys 283
OY 905 -----AAGACGGACGTATCGGTACCCCGAGGACGAGCAAGAAATCCCAAG 952
DB 284 AspIleTyrProGluMetAlaGserIleLysTyrProLysSerGlyThrProAsnProHis 303
OY 953 ATTGCCTTGAACCTGGCTGAGTTCAGACTACAGACGAGGCAAGATGCTGTGACCCAG 1012
DB 304 AlaGluLeuTyrValTyrSerMetLysAspGly----- 314
OY 1013 GAAGAGGAGCTGTGACGCTTCAGCTCGCTGTCCCGAAGGTGAG----- 1060
DB 315 -----ThSerPheHisProArgIleSerGlyAsnLysLys 326
OY 1061 -----TACATCCGACGAGGCGGGGTGACCGGATGAGCAATACGCTGGGCC 1108
DB 327 AspGlySerLeuLeuIleThrGluValIleThrTyrValIleGlyAsnGlyAsn-----ValLeuVal 345
OY 1109 ATGTCTCTGAGCGCGCCGACGAGTGGCTCCACTGCTCTCTCCCGCGCCCTGTTTC 1168
DB 346 LysThrThrAspArgSerSerAspIleLeuThrValPheLeuIleAspThr----- 362
OY 1169 ATGCCGACGACAGAGATGAGGACACCGGCTAGCCTTCGACAGAGTGTGCCAGGAAT 1228
DB 363 IleAlaLysThrSerAsn----- 368
OY 1229 GTCCAGCCGTATGGTGTACGAGAGGTACCAAC-----GTCTGATCAATGTTTCAT 1282
DB 369 -----ValValAlaGAsnGluSerSerAsnGlyLysTyrTyrPrgLuiIleThrHis 384
OY 1283 GACATCTTATCCCTTCCCAATCAGAGGAGAGACGAGCTGTGCTTCTCCGGGCC 1342
DB 385 AsnThrLeu-----PheIleProAla 391
OY 1343 AATGAATGC-----AAGACGGGTCTGCCATTTGTACAAAGTACACCGCGTT 1390
DB 1343 AATGAATGC-----AAGACGGGTCTGCCATTTGTACAAAGTACACCGCGTT 1390
  
```

```

DB 392 AsnGluThrPheAspArgProHisAsnGlyTyr-----ValAspIle 405
OY 1391 TTAAATATCCAGCGCTACAGATTTGAGTGAACCCCTTC---AGCCCGGGGAATGAATT 1447
DB 406 LeuProIleGlyLysTyrAsnHisLeuAlaTyrPheGluAsnSerAsnSerHisTyr 425
OY 1448 AAGTGGCCCATTAAGAAAGATGTGCTGACAGCGGTGAATGGAGGATTTTGGCGAGG 1507
DB 426 Lys-----ThrLeuThrGluGlyLysTyrPrgLuiVal----- 436
OY 1508 CAGCGCTCCAAATCTGGGTCATGAGAGACCAACAGCTGTGACTTCCAGGCGACCAAG 1567
DB 437 AsnGlyProLeuAlaPheAspSerMetGluAsnArgLeu---TyrPheIleSerThrArg 455
OY 1568 GACACCGCGCTGAGACACCACTCTACGTGAGTATGAGCGCGCGCGAGATCGTA 1627
DB 456 LysSerSerThrGluArgHisValTyrTyrIleAspLeuArgSerProAsnGluIleIle 475
OY 1628 CGCCTCACACGCGCGCGC-----TTCTCCATAGCTGTCCATGAGCCAGAAC 1675
DB 476 GluValThrAspThrSerSerGluAspGlyValTyrAspValSerPheSerSerGlyArg 495
OY 1676 TTGCATATGTTCTGACACCACTACAGCAGGTGAGCAGCGCGCTGCGTG-----CAC 1729
DB 496 PheGlyLeuLeuThrTyrTyrLysGlyProLysValProTyrGluLysIleValAspPheHis 515
OY 1730 GTCTACAGCTGAGCGCGCGCGCGAGCAGACCCCTGACACAGCGCGCGCTTGGGCT 1789
DB 516 SerArgLysAlaGluLysCysAspLysGlyAsnValLeuGlyLysSerLeuTyrHisLeu 535
OY 1790 AGCATGATGAGGACA---GCCACGTGCGCGCGCGATTAAT---GTCTCCACAGATCTTC 1843
DB 536 GluLysAsnGluValLeuThrLysIleLeuGluAspTyrAlaValProArgLysSerPhe 555
OY 1844 -----CATTTCCACACGCGCTCGATGCGCGCTCTACGGC 1879
DB 556 ArgGluLeuAsnLeuGlyLysAspGluPheGlyLysAspIleLeuValAsnSerTyrGlu 575
OY 1880 ATGATCTACAAAGCCCGCGCTTCGACCGCGGGAAGAACACCCCGCTTCCTTGTA 1939
DB 576 IleLeuProAsnAspPheAspArgLuiThrLeuSerAspHisTyrProValPhePhePheAla 595
OY 1940 TATGAGGCGCGCGAGTGCAGCTGTGAATACCTCTTCAA---GCCATCAAGTACTTC 1996
DB 596 TyrGlyGlyProAsnSerGlnGluValIleValLysThrPheSerValGlyPheAsnGluVal 615
OY 1997 CGGCTCAACACACTGCGCTCCCTCGCGGTACGCGCTGTTGTGATTGACGCGAGGCGCTCC 2056
DB 616 -----ValAlaSerGlnLeuAsnAlaIleValValValAspGlyArgGlyThr 632
OY 2057 TGTACAGGAGGCGTTCGGTTCGAAGGCGCGCTGAATAACCAATGGCGCAGTGAATC 2116
DB 633 GlyPheLysGlyGlnAspPheArgSerLeuValAlaArgAspArgLeuGlyAspTyrGluAla 652
OY 2117 GAGGACCAAGTGAAGGCGCTGACGTTCTGTGCGCCAGAAATATGAGC-----TTCAATC 2167
DB 653 ArgAspIleIleSer-----AlaAlaSerLeuTyrGlySerLeuThrPheVal 668
OY 2168 GACCTGACCGGATGTCATCCATGCTGCTGCTACGCGGCGCTTCCTGCTCATAGGGG 2227
DB 669 AspProGlnLysIleSerLeuPheGlyTyrPserTyrGlyLysIleuThrLeuLysThr 688
OY 2228 CTAAATCCACAAG---CCCCAGGTGTTCAAGGTGGCCATCCCGGGTGGCCGCTACCGTTC 2284
DB 689 LeuGluLysAspGlyGlyAlaGlnHisPheLysTyrGlyMetSerValAlaProValThrAsp 708
OY 2285 TGATGGCTTACACACAGGATACACTGACGCTACATGAGACGCTCCCTGAACAACAACAG 2344
DB 709 TyrArgPheTyrAspSerValTyrThrGluArgTyrMetHisThrProGlnGluAsnPhe 728
OY 2345 CACGGCTATGAGGCGGTTCCGCGCGCGCTGCACGTGAGAGAAAGTGGCCATGAGCCCAAC 2404
DB 729 AspGlyTyrValGluSerSerVal-----HisAsnValThrAlaLeuAlaGlnAlaAsn 746
  
```

Db 227 SerAspTrpLeuTyrGlu-----GluGluIle 235
 QY 758 TTGCAGCGCTTCACGTGGTACTGGTGGGCGCCACAGCGCTCTGGGAGATTCCAGAGGC 817
 Db 236 LeuLysThrHisIleAlaHisTrpSerProAspGlyThr----- 249
 QY 818 CTCAGAGCGTGGCAATCTGTATGAGAGAACTGCATGATCCGAGAGTGGAGTCAATTCAC 877
 Db 250 -----ArgLeuAlaTyrAlaLeuThrIleAsnAspSerArgValProValMetGlu 265
 QY 878 GTCCCTCTCTCGCGCTAGAAAGAAAGAGACG---GACTGTATCGGTACCCAGAGCA 934
 Db 266 LeuProThrTyrThrGlySerValTyrProThrAlaLysProGlyHisTyrProLysAla 285
 QY 935 GGCAGCAAGAAATCCCAAGATTGCGCTGAACCTGGCTGATGCCAGACATGACAGCCAGGC 994
 Db 286 GlyLysGluAsnProSerIleSerLeuHisValIle-----Gly 298
 QY 995 AAGATCGTCTCGACCCAGAGAGAGAGAGCTGGACCGCTTCAGCTCGCTGGTCCGAG 1054
 Db 299 LeuAsnGlyProThrHisAspLeuGluMetThrProProAspArg-----ProArg 315
 QY 1055 GTGGAG-----TACATCGCCAGGCGCGGTGACCCCGGATGGCAATACGCTGGGCC 1108
 Db 316 MetArgGluTyrTyrIleThrMetValLysTrpAlaThrSerThrLysValAla---Val 334
 QY 1109 ATGTTCTGGAGCGGCGCCACAGTGGCTCAGCTGCTCTCTCCCGCGCGCTGTC 1168
 Db 335 AsnTrpLeuSerArgAlaGlnAsnValSerIleLeuThrLysCysAspAlaThrThrGly 354
 QY 1169 ATCCCGAGCAGCAAGAAATGAGAGACAGCGGCTGACCTCTGCCAGAGCTGCCAGAGAT 1228
 Db 355 ValLysThrLysLysHisGluAspGlu----- 363
 QY 1229 GTCCAGCCGATGTGGTGTACAGAGAGGTCAACAGCTGGATC-----AATGTT 1279
 Db 364 -----SerGluAlaTrpLeuHisArgGluAsnGlu 373
 QY 1280 CATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTGCTTCTCTCCG 1339
 Db 374 GluProValPhe-----SerLysAspGlyArgLysPhePhePheValArg 388
 QY 1340 GCCAATGATGCAAGACCGGCTTCTGCATTGTACAAGTCAACCGCGCTTTAAATCC 1399
 Db 389 AlaIlePro---GlnGlyGlnGlnGlyLysPheThrHisIleThrVal----- 403
 QY 1400 CAGGGCTACGATTGGAGTACGCTTCACCGCGGGGAAGATTAATTAGTCCCAT 1459
 Db 404 -----SerSerSerGlnProAsnSerSerAsnAspAsn-----Ile 415
 QY 1460 AAGGAGAGATGCTGTGACCGAGCGGGAATGGAGGTTTGGGAGGACAGGCTCCAG 1519
 Db 416 Gln-----SerIleThrSerGlyAspTrpAspVal-----ThrLys 427
 QY 1520 ATCTGG---GTCAATGAGAGAACCAACTGTGTACTTCCAGGACACCAAGACAGCGCG 1576
 Db 428 IleLeuSerTyrAspGlyLysArgSerGlnIleTyrPheLeuSerThrGluAspLeuPro 447
 QY 1577 CTGGAGACCACTCTACGTGGTCAAGCTATGAGCGCGCGGCGGAGATCTGATCGCTCAC 1636
 Db 448 ArgArgArgGlnLeuTyrSerAlaSer-----Thr 457
 QY 1637 ACGCGCGGCTTTCCTCCATAGCTGTCCATGACGCAAACTTCGACATTTCTGTACGCCAC 1696
 Db 458 ValGlySerPheAsnArgGlnCysLeuSerCysAspLeuValAspAsnCysThrTyrPhe 477
 QY 1697 TACAGCAGCGTACGACGCGCGCTGCGTACCTTACAGCTGAGAGCGCCCGAC--- 1753
 Db 478 SerAlaSerPheSerProGlyAlaAspPheLeuLeuLysCysGluGlyProGlyVal 497
 QY 1754 -----GACGACCGCTTCACACAGCAGCGCGCGCTTCTGGGCTAGCATGATG 1798
 Db 498 ProThrValSerValHisAsnThrThrAspLysLysMetPheAspLeuGluThrAsn 517

QY 1799 GAG---GCAGCAGAGTCCCGCCCGGATTAATGTCTCCAGAGATC---TTCATTTTCCAC 1852
 Db 518 GlnHisValGlnLysAlaIleSerAspArgGlnMetProLysValGluTyrArgLysIle 537
 QY 1853 ACGGCTCGATGTCCGCGCTTACGCGCATGATTCACAGCCCGCCAGCTTCCAGCGAGG 1912
 Db 538 GluThrAspAspTyrAsnLeuProIleGlnIleLeuLysProIleThrPheThrAspThr 557
 QY 1913 AAGAAGCACCCCGCTCTTTGTATATGAGAGGCGCCCGAGGTGCACTGGATATAC 1972
 Db 558 AlaHisTyrProLeuLeuLeuValValAspGlyThrProGlySerGlnSerValAlaGlu 577
 QY 1973 TCTTCAAGGCAATCAATCACTTGGCTGGCTCAACACAGCTGGCTGGCTGGCTGGCTGG 2032
 Db 578 LysPhe---AlaValThrTrpGluThrVal---MetValSerSerHisGlyAlaValAl 595
 QY 2033 GTTGTGATTCAGCAGCAGCGCTCTGTGTCAGCAGAGGCTTCCGTTTCAAGGCGCTGAA 2092
 Db 596 ValLysCysAspArgLysArgGlySerGlyPheGlnGlyThrArgLeuLeuHisGluValArg 615
 QY 2093 AACCAATGGCCCAAGTGCAGATGAGACGACGAGTGGAGGCGCTGCACTTGGCGGAG 2152
 Db 616 ArgArgLeuGlySerLeuGluGluLysAspGlnMetGluAlaValArg---ValMetLeu 634
 QY 2153 AAGTATGCTCATTCAGCAGCTGAGCGGAGTGGCAATCCATGCTGCTGCTGCTGCTGCTG 2212
 Db 635 LysGluProTyrIleAspLysThrArgValAlaValPheGlyLysAspTyrGlyLys 654
 QY 2213 CTCTCGCTCATAGGCGCTTAATCCAGAG-----CCCGAGGTGTTCAAGGTGGCC 2260
 Db 655 LeuSerThrTyrLeuLeuProAlaLysGluLysArgGluAlaProValPheSerCysGly 674
 QY 2261 ATCGCGGTGCGCGGTACCGCTGCGATGCGCTTACACACAGAGTCACTGAGCGCTAC 2320
 Db 675 SerAlaLeuSerProIleThrAspPheLysLeuTyrAlaSerAlaPheSerGluArgTyr 694
 QY 2321 ATGGACGCTCCGAGAACACACAGCAGCGCTATGGAGCGGCTGCGCGCTGCGAGCTG 2380
 Db 695 LeuGlyLeuHisGlyLeuAspAsnArgAlaTyrGluMetAlaLysValAlaHisArgVal 714
 QY 2381 GAGAGCTGCCCAATGAGCCCGCCAGCTGTATCTCTCCAGCGCTTCCAGGAGAAAC 2440
 Db 715 SerAlaLeu-----GlnGlyGlnGlnPheLeuValIleHisAlaThrAlaAspGluLys 732
 QY 2441 GTGACATTTTTCACACAACTCTCTGCTCCCACTGATCCGACGAGGAAACCTTAC 2500
 Db 733 IleHisPheGlnHisThrAlaGluLeuIleThrGlnLeuIleLysGlyLysAlaAsnTyr 752
 QY 2501 CAGCTCGATCTACCCCAACAGAGACAGATTAATGCTGCTGCGCGCGGAGAGCAG 2560
 Db 753 SerLeuGlnIleTyrProAspGlnSerHisTyrPheSerSerAlaAlaLeuGlnGlnHis 772
 QY 2561 TATGAGTCACTGTCTGCTCACTTCTTACAGGAA 2593
 Db 773 LeuHisArgSerIleLeuGlyPhePheValGlu 783
 RESULT 14
 A30107
 dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YNR028C
 C:Species: Saccharomyces cerevisiae
 C:Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
 C:Accession: S46780; A30107
 R:Du, Z.
 submitted to the EMBL Data Library, June 1994
 A:Description: The sequence of S. cerevisiae cosmid 8082.
 A:Reference number: S46773
 A:Accession: S46780
 A:Molecule type: DNA
 A:Residues: 1-818 <DDZ>
 A:Cross-references: EMBL:U10399; NID:9500689; PIDD:AB68879.1; PIDD:9500698; GSPDB:GNO
 R:Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.

```

Db 350 AspaIaThrThnGlyAlcLysThrLysHisGluAspGlu----- 363
QY 1214 GGTGTCACAGAGATGTCACGCGTATGTGTGTACAGAGAGTCCACAACTGTGATC 1273
Db 364 -----SerGluAlaTrpLeu----- 368
QY 1274 -----AATGTCATGACATCTTATCCCTCCCTCCCAATCAGAGGAGAGACGAG 1324
Db 369 HisArgGlnAsnGluIuProValPhe-----SerLysAspGlyArgLys 383
QY 1325 CTCTGCTTTCCTCCGCGCCCATGATGATGACAGACGCGTCTGCCATTTGTACAAAGTCAC 1384
Db 384 PhePhePheIleArgValIlePro---GlnGlyGlyArgGlyLysPhePheYrHisIleThr 402
QY 1385 GCGGTTTAAATCCACAGGCTACGATGTGAGGACCGCTCCGCGCGCGGAGATGAA 1444
Db 403 Val-----SerSerGlnProAsnSerSerAsnAspAsn--- 414
QY 1445 TTTAACTGCCCATTAAGAGAGATGTTGCTGACACGCGGTGAATGGAGGTTTGGCG 1504
Db 415 -----IleGln-----SerIleThrSerGlyAspTrpAspVal----- 425
QY 1505 AGGACGCGCTCCACAGATCTGGGTC---AATGAGAGACCAAGCTGCTACTTCCAGGCG 1561
Db 426 -----ThrLysIleLeuAlaTyrAspGluLysGlyAsnLysIleTyrPheLeuSer 442
QY 1562 ACCAAGACACGCGCGCTGGAGCACGCTCTACGTGTCAGTATGAGGCGCGCGCGAG 1621
Db 443 ThrGlnAspLeuProArgArgArgGlnLeuTyrSerAlaAsnThrGluGly----- 459
QY 1622 ATGCTAGCGCTCACACAGCGCGCGCTTCCCAT-AGCTGCTCATGAGCCAG 1672
Db 460 -----AsnPheAsnArgGlnLysLeuSerGlyAspLeuValGlu 472
QY 1673 AACTTGACATGTTCTGACACCCCTACAGCAGCGGTGACAGCGCGCGCTTCGTCGACGTC 1732
Db 473 AsnGlyThrTyrPheSerIleAspSerPheSerHisSerMetAsp-----PhePheLeu 489
QY 1733 TACAACTGACGCGCGCGCGCGC-----GACGACCGCGCTGACAGACGAG 1774
Db 490 LeuLysCysGluGlyProGlyValAlaProMetValThrValHisAsnThrAspLysLys 509
QY 1775 CCGCGCTTGGGCTAGCATGATGAG---GCAGCAGCTGCGCGCGCGATTAATGTTCC 1831
Db 510 LysMetPheAspLeuGluThrAsnGlnHisValLysAlaIleAsnAspArgGlnMet 529
QY 1832 CCAGACATCTTCATTCACACAGCGC---TCGATGTCGCGCTACAGCATGATCTAC 1888
Db 530 ProLysValGluTyrArgAspIleGluIleAspAspLysAsnLeuProMetGlnIleLeu 549
QY 1889 AAGCCACGCGCTGACGACGAGGAGAAAGCAGCCGCTCTGTTGATATGAGGCG 1948
Db 550 LysProAlaThrPheThrAspThrThrHisTyrProLeuLeuValValAspGlyThr 569
QY 1949 CCCCAGGTGACGCTGATTAATCTCTCAAGGACATCACTAGCTGCGCTCAACACA 2008
Db 570 ProGlySerGlnSerValAlaGluLysPheGlu---ValSerTrpGluThrVal---Met 587
QY 2009 CTGGCTCCCTGGGCTACGCGCGCTGTGATGACGCGCGCGCTCTGTCACGAGG 2068
Db 588 ValSerSerHisGlyAlaValValLysCysAspGlyArgGlySerGlyPheGlnGly 607
QY 2069 CTTCGCTTCCGAAAGGCGCGCGGAAACCAATGCGCAGGTGAGATCGAGACGAGT 2128
Db 608 ThrLysLeuLeuHisGluValAlaArgArgArgLeuGlyLeuLeuGluGlyAspGlnMet 627
QY 2129 GAGGCGCTCAGTCTGCGCGCGAGATGATGCTCATGCACTGACCGCAGTTCAT 2188
Db 628 GluAlaValAlaArgThrMetLeu---LysGluGlnTyrIleAspArgThrArgValAlaVal 646
QY 2189 CATGCTGCTCCTACGCGCGCTCTCTGCTCATGCGGCTAATCCACAAG----- 2239
Db 647 PheGlyLysAspTyrGlyGlyTyrLeuSerThrTyrIleLeuProAlaLysGlyGluAsn 666

```

```

QY 2240 ---CCGACGTGTTCAGGTGGCATCGCGGTGCCCGGTCACGCTGTGATGGCTAC 2296
Db 667 GlnGlyGlnThrPheThrCysGlySerAlaLeuSerProIleThrAspPheLysLeuTyr 686
QY 2297 GACACAGGTATACAGCGCTACATGACGTCCCTGAGAACACGACGACGCTATGAG 2356
Db 687 AlaSerAlaPheSerGluArgTyrLeuGlnHisGlyLeuAspAsnArgAlaTyrGln 706
QY 2357 GCGGTTCCGTCGCGCTGACGCGGAGAGCTGCCAATGAGCCCAACCGCTTGTATC 2416
Db 707 MetThrLysValAlaHisArgValSerAlaLeu-----GluGluGlnGlnPheLeuIle 724
QY 2417 CTTCACGCGCTTCCGCGCAAAACGTCATTTTCCACAAACTCTGCTGCCAA 2476
Db 725 IleHisProThrAlaAspLeuLysIleHisPheGlnHisThrAlaGluLeuIleThrGln 744
QY 2477 CTGATCCGACGAGGAAACCTTACAGCTCCAGATCTACCCCAACGACAGACAGTAT 2536
Db 745 LeuIleArgGlyLysAlaAsnTyrSerLeuGlnIleTyrProAspLeuSerHisTyrPhe 764
QY 2537 CGCTGCCCGAGTCGCGCGGACGACATGAACTGCTGCTGCTTGTACAGGAA 2593
Db 765 ThrSerSerLeuLysGlnHisLeuTyrArgSerIleLeuAsnPheValGlu 783

RESULT 13
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C/species: Bos primigenius taurus (cattle)
C/date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C/accession: A41793
R/Wade, K.; Yokotani, N.; Hunter, C.; Dol, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A/title: Differential expression of two distinct forms of mRNA encoding members of a
A/reference number: A41793; M01D:92108018; PMID:1729689
A/status: preliminary; not compared with conceptual translation
A/molecule type: mRNA
A/residues: 1-803 <MAD>
A/cross-references: GB:M76429; NTD:9408719; PIDN:A041623.1; PID:9408720
A/notes: sequence extracted from NCBI backbone (NCBIP:75138)
C/superfamily: dipeptidyl-peptidase IV
C/keywords: dipeptidylpeptide hydrolase; glycoprotein
F.257/342/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 1,06e-18 Length: 803
Score: 442.50 Matches: 177
Percent Similarity: 38.99% Conservative: 108
Best Local Similarity: 24.21% Mismatches: 287
Query Match: 9.23% Indels: 159
DB: 2 Gaps: 28

US-09-976-674-4 (1-2617) x A41793 (1-803)
QY 458 CTCTCTCTTCCAGGCGCAGACAGCTCTTCCATGTCGCGAGCGGCGCAAGAGCG 517
Db 193 IlePheIlePheGlu-----AsnAsnIleTyrTyrCysAlaHisValGlyLysGlnAla 210
QY 518 TTTCATGTGTCCCTATGAACCGCTGGAATCAAGACCCAGTGCAGGCGCGCGATG 577
Db 211 IleArgVal----- 213
QY 578 GACCCCAAAATCTGCCCTGCGGACCTGCTTCTTCTTCATCAATAACGACACTG 637
Db 213 ----- 213
QY 638 TGAGTGCCAAATCAGACAGAGCGAGCGCGGCTGACCTTTCGACCAAGGTTTA 697
Db 214 -----ValSerThrGlyLysGluGlyValIle-----TyrAsnGlyLeu 226
QY 698 TCCAAATGCTCGTGGATGCCCAAGACTGCGGCGTGGCCACCTTGCTCATACAGGAAG 757
Db ----- 757

```

QY 1655 -ACGTGCTCATGAGCCAGACATTCGATGTCAGCCACTACAGACGCTGAGCAC 1713
 ||||| : : : : :
 Db 528 uSerCysAspLeuValGluAsnCysThrTyrPheSerAlaSerPheSerHisSerMetAs 548
 QY 1714 GCCGCCCTGGGTGCAGCTCTACAAAGCTGACGGCCCCGAC-----GA 1755
 : : : : :
 Db 548 p-----PhePheLeuLeuLysCysGluGlyProGlyValProMetValThrValH 565
 QY 1756 CGACCCCTGTCAGACAGACGCCCTCTGCGGTAGCATGATGAG---GCAGCCACTG 1812
 : : : : :
 Db 565 sAnThrThrAspLysLysLysMetPheAspLeuGluThrAsnGluHisValLysLysAl 585
 QY 1813 CCCCCGGATTATGTCCTCCACAGATTCCTCCATTCACACGCGC---TCGGATTGCG 1869
 ||| : : : : :
 Db 585 AlLeAsnAspArgGlnMetProLysValGluTyrArgAspIleGluLeuLeuAspLys 605
 QY 1870 GCTCTACGGCATATCTACAAAGCCCCAGCCTTGACCCAGGAGAGAACACCCACCGT 1929
 ||| : : : : :
 Db 605 nLeuProMetGlnIleLeuLysProAlaThrPheThrAspThrThrHisTyrProLeu 625
 QY 1930 CCTCTTTGATATGAGGCCCCGAGTGCAGCTGGTGAATTAATCTCTTCAAGGATCA 1989
 : : : : :
 Db 625 uLeuValAlaAspGlyThrProGlySerGlnSerValAlaGluLysPheGlu---ValSe 644
 QY 1990 GTACTTGGCGCTACACACACTGCGCTCCCTGGCTACGCGCTGCTGATTCAGCGGAG 2049
 : : : : :
 Db 644 rTTPGluThrVal---MetValSerSerHisGlyAlaValAlaValLysCysAspGlyAr 663
 QY 2050 GGGCTCCTGTACAGCAGAGGCTTGGTTGAGAGGGCCCCGAAACCAATGGCGCAGT 2109
 ||||| : : : : :
 Db 663 gGlySerGlyPheGlnGlyThrLysLeuHisGluValAlaArgArgLysLeuLeu 683
 QY 2110 GGAGATGAGGACGAGGCTGGAGGCGCTGAGTGGCCGAGAGTATGGCTTACAGCA 2169
 : : : : :
 Db 683 uGluGluLysAspGlnMetGluAlaValAlaArgThrMetLeu---LysGluGlnTyrLeu 702
 QY 2170 CCTGAGCCGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2229
 : : : : :
 Db 702 pArgThrArgValAlaValPheGlyLysAspLysGlyLysLysSerThrTyrIleLe 722
 QY 2230 AATCCACAG-----CCCCAGGTTCGAAAGTGGCCATGCGGGGCTCCCGGT 2277
 ||| : : : : :
 Db 722 uProAlaLysGlyLysAsnGlnGlyThrPheThrCysGlySerAlaLeuSerProI 742
 QY 2278 CACGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2337
 : : : : :
 Db 742 eThrAspPheLysLeuTyrAlaSerAlaPheSerGluArgTyrLeuGlyLeuHisGlyLe 762
 QY 2338 CAACGAGCAGGCTATGAGCGGGTCCGTGCGCTGACAGTGAAGAGTGCCTGCAATGA 2397
 : : : : :
 Db 762 uAspAsnArgAlaTyrGlnMetThrLysValAlaHisArgValSerAlaLeu-----G 780
 QY 2398 GCCCAACCGCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2457
 : : : : :
 Db 780 uGluGlnGlnPheLeuIleLeuHisProThrAlaAspGlyLysIleHisPheGlnHis 800
 QY 2458 AAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2517
 : : : : :
 Db 800 rAlaGluLeuIleThrGlnLeuIleArgGlyLysAlaAsnTyrSerLeuGlnIleTyr 820
 QY 2518 CAACGAGCAGCAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2577
 : : : : :
 Db 820 oAspGluSerHisTyrPheThrSerSerSerLeuLysGlnHisLeuTyrArgSerIleI 840
 QY 2578 GCACCTTCTACAGGAA 2593
 : : : : :
 Db 840 eAsnPhePheValGlu 845
 RESULT 12
 168600
 dipeptidyl aminopeptidase like protein - human
 C:Species: Homo sapiens (man)

C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
 C/Accession: I68600
 R/Yokotani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
 Hum. Mol. Genet. 2, 1037-1039, 1993
 A>Title: Non conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-re
 A/Reference number: I54331; MUID:9372805; PMID:8103397
 A/Accession: I68600
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-803 <RES>
 A/Cross-references: GB:I68600; NID:9306707; PIDN:AAA3761.1; PID:9306708
 C/Superfamily: dipeptidyl-peptidase IV

 Alignment Scores:
 Pred. No.: 6, 99e-19 Length: 803
 Score: 445.50 Matches: 176
 Percent Similarity: 40.05% Conservative: 120
 Best Local Similarity: 23.82% Mismatches: 268
 Query Match: 9.29% Indels: 175
 DB: 2 Gaps: 31

 US-09-976-674-4 (1-2617) x I68600 (1-803)
 QY 458 CTTCTCTTCCAGGCGCAGACAGCCTCTCTCATGTGCGCAGCGCGCAGAACGGC 517
 : : : : :
 Db 193 lIePheIlePheGlu-----AsnAsnIleTyrTyrCysAlaHisValGlyLysGlnAla 210
 QY 518 TTCATGTGTTCCTTATGAAACCGCTGGAATCAAGACCCAGTGTCTCAGGCCCCGATG 577
 ||| : : : : :
 Db 211 lIeArgVal----- 213
 QY 578 GACCCCAAAATGTCCTGCGCAGCCTGCTTCTCTTCAATTAACAGGACCTG 637
 : : : : :
 Db 213 ----- 213
 QY 638 TGGGTGGCAACATCGAGACAGCGGAGCGCGGCTGCTTGCACCAAGTTTA 697
 : : : : :
 Db 214 -----ValSerThrGlyLysGluGlyValIle-----TyrAsnGlyLeu 226
 QY 698 TCCAAATCTCTGATGAGACCCCAAGTGTGCGGCTGTGCGCACCTTGTCTATACAGAAAG 757
 ||||| : : : : :
 Db 227 SerAspTrpLeuTyrGlu----- 235
 QY 758 TTCAGCGCTTCACTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 817
 : : : : :
 Db 236 LeuLysThrHisIleAlaHisTyrTrpSerProAspGlyThr----- 249
 QY 818 CTCAGACGCTGCGAATCTGTATAGAGAAATCGATGATCGAGTGAAGTCAATTGAC 877
 : : : : :
 Db 250 -----ArgLeuAlaTyrAlaAlaIleAsnAspSerArgValProIleMetGlu 265
 QY 878 GTCCCTCT-----CTGCGCTGAAGAAAGAAAGAGAGAGCTGAT 919
 : : : : :
 Db 266 LeuProThrTyrThrLysSerIleTyrProThrValLys-----ProTyr 280
 QY 920 CGGTACCCCAAGCAGCAGCAAGAAATCCCAAGATTGCTCTGAAACTGGCTGAGTTCCAG 979
 : : : : :
 Db 281 HisTyrProLysAlaGlySerGluAsnProSerIleSerLeuHisValIle----- 297
 QY 980 ACTGACAGCCAGGCAAGATGCTGTGACCCAGAGAGAGAGTGTGTGAGCCCTTCAGC 1039
 : : : : :
 Db 298 -----GlyLeuAsnGlyProThrHisAspLeuIleMetMetProProAspAsp 313
 QY 1040 TCCTGTCTCCGAAAGTGGAG-----TACATGCGCAGGCGCGGCTGAGCCCGGATGCG 1093
 : : : : :
 Db 314 -----ProArgMetAlaGlyTyrTyrIleThrMetValLysTrpAlaThrSerThr 330
 QY 1094 AAATPAGCCTGGGCGCATGTGCTGAGCGCGCGCAGCAGTGTCCAGTGTGCTGCTGCTGCTG 1153
 : : : : :
 Db 331 LysValAla---ValThrTrpLeuAsnArgAlaGlnAsnValSerIleLeuThrLeuCys 349
 QY 1154 CCCCCGCGCCTGTTCATCCCGAGCAGAGAAATGAGAGAGCGGCTAGCCTGTGCGAGA 1213
 : : : : :


```

RESULT 11
154331
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I54331
R:Yokotani, N.; Doi, K.; Wenthoid, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I54331
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-865 <RES>
A:Cross-references: GB:M96859; NID:9306705; PIDN:AAA35760.1; PID:9306706
C:Superfamily: dipeptidyl-peptidase IV

Alignment Scores:
Pred. No.: 2,93e-19 Length: 865
Score: 452.00 Matches: 221
Percent Similarity: 39.96% Conservative: 149
Best Local Similarity: 23.87% Mismatches: 369
Query Match: 9.43% Indels: 187
DB: Gaps: 43

US-09-976-674-4 (1-2617) x I54331 (1-865)
OY 76 CGCGTCCAGGTGACAGACACTGTGGAGGCTCGGAGCATCAT---CCACGGCAG 132
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 21 ProProGluAlaSerHisLeuLeuGlyGlyGlnGlyProGluGlnAspGlyGlyAlaGly 40
OY 133 CGCAGACTCTCGGCGCTCATTTGTCACAAGCGCCCGCCACGACTTCCAGTTTGTGCAGA 192
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 41 AlAlyProLeuGlyProAlaGlnAlaAlaProArg-----Glu 55
OY 193 GAGCGATGAGTGGGCGCCACATCC-----CACCGCTTACTACTGGGAATGCC 243
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 56 ArgGly-GlyGlyGlyGlyGlyGlyGlyGlyGlyArgProArgPheGlnTyrGlnIArgse 75
OY 244 ATATGGCAGCGGAGCAACACTCTCTCTACTGTAGATCCCAAGAGGTCGGAGA 303
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 75 rAspGlyAspGluGlnAspGluLeuValGlySerAsnProProGlnArgAsnTrpIysG 95
OY 304 GGCTCTGCTCTCTCTCTCTCTGTAAGACATGTGATCATTTCCAGGCCAGCCACCA 363
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 95 yLeuAlaAlaLeuValAlaLeuValIleCysSerLeuIleValThrSerValIl 115
OY 364 TGGGGTCTACTCTCGGAGGAGAGAGCTGCTGAGAGCGGAAAGCGCTGGGGTCTTCGG 423
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 115 eLeuLeuThrProAlaGlnAspAsnSerLeuSerGlnIlyLysValThrValGluAs 135
OY 424 CATCACTCTCTACGACTTC-----CACAGCGAGAGTGGCTCTCTCTCTCTCTCTCT 477
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 135 pLeuPheSerGlnAspPheIysIleHisAspProGluAlaLysTrpIle-----Se 152
OY 478 CAACAGCCTCTCCACTGTCGAC-----502
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 152 rAspThrGlnPheIleTyrArgGlnIlyGlnGlyThrValArgLeuThrAsnValGluTh 172
OY 503 -----GGCGGCAAGAAC-----GGCTT 519
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 172 rAsnThrSerThrValLeuIleGluGlnIlyLysIleGlnSerLeuArgAlaIleArgTy 192
OY 520 CATGATGTCCCT-----ATGAACCGCTGGAAT 549
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 192 rGluIleSerProAspArgGlnIlyThrAlaLeuPheSerTyrAsnValGlnIle----Ty 211
OY 550 CAAGACCACTGCTCAGGGCCCGGATGAGCCCAAAATGCTGCCCTGCCAGCTGCTT 609
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 211 rGlnHisSerTyrThrGlyValLeuSerIleLeuProHisGlnAspProGlnse 231
OY 610 CTTCCTCC-----TTCATCAATAAGACGACCTGTGGTGGCCAAATC---GAGACAG 660
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db 231 rLeuAspProProGluValSerAsnAlaLysLeuGlnIlyThrAlaGlyTrpIlyProIysG 251
OY 661 CGAG-----GAGCGCGGCTGACCTTCTGCTC---CACCAAGATTATC 699
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 251 yGlnGlnLeuIlePheIlePheGlnAsnIleIlyrTyCysAlaHisValGlyLysG 271
OY 700 C-----AATGTCCTGATGATGCCCAAGTCTGCG-----GGTGGCCACTT 741
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 271 nAlaIleArgValIleSerThrGlyGlnIlyValIleTyrAsnGlyLeuSerAspTr 291
OY 742 CGTCATACAGGAAGAG---TTGGACCGCTTCACTGGGTACTGGTGGTCCCAACAGCTC 798
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 291 pLeuTyGlnGlnGlnIleLeuLysThrHisIleAlaIleStrTrpSerProAspIlyTh 311
OY 799 CTGGGAAGTTACAGAGGCGCTCAACAGCGTGGCAATCTGTATGAGAGAGTGGATGAGTC 858
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 311 r-----ArgLeuAlaIlyrAlaAlaIleAsnAspse 321
OY 859 CGAGGTGAGGTCAATTACGTCCTCT-----CTGCGCTGACAGA 900
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 321 rArgValProIleMetGlnLeuProThrTyrThrGlySerIleTyrProThrVallys-- 340
OY 901 AAGGAAGACGAGCTGTATCGGTACCCCGACAGCAGCAGCAAGATCCCAAGATTGCTT 960
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 341 -----ProTyrHisTyrProIlyAspIlySerGlnAsnProSerIleSerIle 356
OY 961 GAACAGCTGAGATTCCAACTGACAGCCAGGCAAGATGCTGTGACCCAGAGAGA 1020
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 356 uHisValIle-----GlyLeuAsnGlyProThrHisAspLeuI 369
OY 1021 GCTGTGACAGCCCTTCACTGCTGCTTCCCGAAGTGGAG-----TACATGCCAGGCC 1074
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 369 uMetMetProProAspAsp-----ProArgMetArgIlyTyrIleThrMetVa 386
OY 1075 CGGGTGACCCGGGATGAGCAATACGCTGGGCGCATGTCCTGACCGCCCGCCACAGTG 1134
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 386 lLysTrpAlaThrSerThrLysValAla---ValThrPheAsnAlaArgAlaGlnsVa 405
OY 1135 GCTCCAGCTGCTGCTCTCTCCCGCGCCCTGTTCATCCGAGACAGCAATGAGAGCA 1194
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 405 lSerIleLeuThrLeuCysAspAlaThrThrGlyValCysThrLysLysHisGlnAsp 425
OY 1195 GCGGTAGCCTTGCCAGAGCTGTCCCAAGAAATGTCAGCCGTATGTGTACAGAGA 1254
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 425 u-----425
OY 1255 GGTCAACAACGTGTGATC-----AATGTTCATGACATCTTATCCCTTCCCCA 1305
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 426 ---SerGluAlaThrPheHisArgGlnAsnGlnIlyProValPhe-----439
OY 1306 ATCAGAGGAGAGAGAGAGCTGTCTTCCGCGCCAAATGATGCAAGACGCGCTTGTG 1365
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 440 -SerLysAspGlyArgLysPhePheIleArgAlaIlePro---GlnGlyLysArg 458
OY 1366 CCAATTGTACAAAGTCACGCGCGTTTAAATCCACAGGCGTCAGATTGAGTGAGCCCTT 1425
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 458 yLysPheTyrHisIleThrVal-----SerSerSerIleProAs 471
OY 1426 CAGCCCGGGGAAGATGAATTAAGTCCCATTAAGAGAAGATGCTGTGACGCGG 1485
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 471 nSerSerAsnAspAsn-----IleGln-----SerIleThrSer 483
OY 1486 TGAATGAGAGGTTTGGCGAGCAGCGCTCCAAAGATCTGGGCT---AATGAGAGACCA 1542
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 483 yAspTrpAspVal-----ThrLysIleLeuAlaIlyrAspGlnIlyLys 498
OY 1543 GCTGTGTACTTCCAGGCAACAGACAGCGCGCTGAGACACACCTTACCTGTGTCAG 1602
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 498 nLysIleTyrPheLeuSerThrGlnLysPheuProArgArgArgGlnLeuTyrSerAlaAs 518
OY 1603 CTATGAGCGCGCGGAGAGATGCTACGCGCTCACACAGCGCGCGCTTCCCAT----- 1654
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 518 nThrGlnGly-----AsnPheAsnArgGlnCysIle 528

```

```
Db 142 sleuylsilevalasnansn-----gluargil 152
OY 553 GACCCAGCTGCATCAGGCGCCGATGAGACCCAAAATCTGCCCTGCC-----GAGCC 603
Db 152 ethrtyraspilleuylleuylgslugluservalilleglinalphelystrpasngl 172
OY 604 TGCCTTC-----TTCCTTCATCAATTAACAGCGACCTGTGGGTGGCCAAATCGAGAC 657
Db 172 ylyspheasnaphavalphevalgluserasnlyslletryrclnser-----se 190
OY 658 AGCGAGAGACGGGGGTGACCTTCCTGCACCAAGTTTATCCATGCTCGGATGACCC 717
Db 190 rproglinalserlyleuylleuylrargvalserasnlygllygluylsthrtylaasp----- 208
OY 718 CAAGTGTGGGTGTGCCACTTCGTATACAGAGAGAG-----TTCAGCGGCTTCAGTGG 774
Db 209 -----Glyleuheasprrpletyrcluglulilehegllyarglyasapal 225
OY 775 GTACTGTGTGCCCCACAGCCTCTGGAGAGTTTCAGAGGCGCTCAAGACGCTGCAGAT 834
Db 225 ametrptrip-----serthlysglyaspqln-----Le 235
OY 835 CCTGTATAGAGAACTCATGATCCGAGGTGAGAGTCATTACGCTCCCTCT---CCTGC 891
Db 235 ualatyrAlaserlyraspsasnlsleuthrlysasnvalseuylsthrtyrhlslar 255
OY 892 GCTAGAGAAAGAGAGAGAGGAC---TGTATTCGTACCCGACGACGACGAGCAGG---AA 945
Db 255 glenugluprottyrprolleasprthrasnphensltyrprolysthrphealalyvalle 275
OY 946 TCCCAAGATTGCCTTGAACCTGGCTGAGTTCACAGCTACGACGACGAGGCAAGATCGTCTC 1005
Db 275 uprothrttyrthleuserlletrpasnlyslsthrclu----- 288
OY 1006 GACCCAGAGAGAGAGCTG---GTCCAGCCCTTCAGCTCCCTGTTCCCGAAGTGAGTA 1062
Db 289 ----GlnserArglInleuaspvalGlnleuylsAspsrleu-----serTyrlhslst 305
OY 1063 CATCGCCAGGCGCGGCTG-----ACCCGGATGGCAATAGCGCTGGC 1107
Db 305 rleuAlaValAllystrleuclulileasnlythrglInleuValserValttrpth 325
OY 1108 CATGTTCCCTGAGCCGCGCCAGCAGTGGTCCAGCTGCCTCCCTCCCGGCGCTGT 1167
Db 325 r-----AsnArgtyrGlnasngluValAlaleuThrllecyasprtrpasrthral 342
OY 1168 CATCCCGAGCAGACAGATAGAGACGCGGCTAGCTTCGCAGAGCTGCCCGAGAA 1227
Db 342 allecysargleuGlnrheglu----- 349
OY 1228 TGTCCAGCCGTATGTGTGTCAGAGAGTCACCAACGTCTGATCAATGTATGATCAT 1287
Db 350 -----tyrlystyrAlaserlysalgttrpa1---Thrhlsaspas 362
OY 1288 CTTCTATCCCTTCCCAATCAGAGGAGAGAGAGCAGCTGTCTGTCTCCCGCCCAATG 1347
Db 362 prhehlslserlle-----ThrsrhegluasprthrleuPheheuleuProhlslas 380
OY 1348 ATSCAAGACCGGCTTTCATTTGTACAAGTACCGCGGTTTAAATCCAGGGCTA 1407
Db 380 ptyrsArgasp-----AsnalrheglInglInvalAlaserleuArg----- 393
OY 1408 CGATTGGAGTACCCCTTCAGCCCCGGGGAAGATGAATTAAGTCCCAATTAAAGAGA 1467
Db 394 -----Leuserhlslsglyln-----LeuArgthtrpr 402
OY 1468 GATTGCTGTGACACGCGGATATGAGGTTTGGCGAGCGACGCGCTCAAGATCTGGGT 1527
Db 402 olusprheleuAsnleucllylutyrsvaltlhrserlleasnly-----Il 418
OY 1528 CAATGAGGAGACCAAGCTGTGACTTCCAGGGCACCAGACGCGCGCTGGAGACCA 1587
Db 418 eaSnlysgluThrArgThrIlePhePhehlslAlaIalalProlysproSerhlslar 438
OY 1588 CCTTCAGTGTGATGATATGAGGCGCGGAGATGTCAGGCTTCACCAAGCGCGGCTT 1647
Db 438 rleuphe-----serlyrserleuAlaaspqln-----serArgsnserlataty 453
OY 1648 CTCCTATAGCTGTCTCATG-----AGCCAGAACTTCGACATGTT 1686
Db 453 rcylslleSerCysserlleuylsasncyslthrtpralaglnlaelInmetklsasprlme 473
OY 1687 CGTCAGCCACTACAGCAGCGGTGAGACGCGCGCGCTGGTGCAC-----GTCTCAAGCT 1740
Db 473 llysthrAlailevalserCyslysglproAlaAlaProhlslthrallevalaenle 493
OY 1741 GAGCGCGCCGACGACGACCCCGCTGCACAAGCAG-----CCCGCGCTTCG 1785
Db 493 uthrArgmetAspsrserlyslsthrGlnhlslasleuThryrAsprlysthrty 513
OY 1786 GCGTACATGATGAGAGCAGCAGCAGCTCCCGGATTAATGTTCTTCAGAGATCTTCA 1845
Db 513 rglInasnArgValgluglInlaegllyleuPro-----VallellelysgluThrlely 531
OY 1846 TTTCCACAGCGCTCGGATGTGCGGCTTCAGCGCATGATCTACAGCCCAAGCCTTGCA 1905
Db 531 slleSerAspAspPheaspAlaleuIlelyslleuSerlle-----ProlyAsprllety 549
OY 1906 GCCAGGAGAGAGCAGC-----CCGACCGTCTTGTATAGGAGGCGCCGACGT 1956
Db 549 rasnalrgAsprlyshlselAlaIleleuIleValhlslValtyrcllyproAsnls 569
OY 1957 GCAAGCTGTGAATTAACCTCTTCAAA---GGCANTCAAGTACTGGCGGCTCAACACATGGC 2013
Db 569 rglInasnThrlysglInlaIlethrGlnlegllyleglu-----gluValAla 585
OY 2014 CTCCTGCGGCTACGCGCTGTTGTG---ATTAGCGAGGCGCTCTGTACGAGGCGCT 2070
Db 585 aserAlaSerGlnAlaIleleuArglleasprlyArglysercllycllyrArglytr 605
OY 2071 TCGGTTCAAGGGGCGCTGAAAAACAAATGGCGAGCTGAGATGAGGACAGCAGTGGGA 2130
Db 605 plystyrArgserAlaIlelyrcllyglInleuGlythrlvalgluValclulnspqlnlely 625
OY 2131 GGGCGCTGCAGTCTGTGCGCCAGAAATAT---GGCTTCATCGACCTGAGCCGAGTGGCAT 2187
Db 625 salallelyvalleuIleuArgleuTyArgnhlslleuIleuAspralarglyAlaIa 645
OY 2188 CCAATGCTGTCTTACGCGGCGCTTCTCTGCTCATGCGGCTAATCCAAAGCC---CA 2244
Db 645 lPhegllytrpserlyrcllyglPheMetThrleuSerMetValasngluAlarproglugl 665
OY 2245 GGTGTTCAAGGTGCGCATGCGGCGGTGCGCGGTCACCGCTGTGGATGGCTACAGACAGG 2304
Db 665 nPhePheLysCysAlaValserValAlarovaltlhrAsnPhelAlatyrtyrAspalath 685
OY 2305 GTACACTGAGCGGTACATG---GACGTCCCTGAGAACCAACAGCAGCGCTATGAGCGGG 2361
Db 685 rtyrthgluArgtyrmetcllyaspralpro-----leuGl 697
OY 2362 TTCGTTGCGCTTCACGTGAGAGCTGCCAAT---GAGCCCAACCGTTCCTATCTCT 2418
Db 697 usertyrserAspvaltlhrlyslsleuAspAsnPhelyserThrArgleuLeuIeume 717
OY 2419 CCAGCGCTTCCTGAGGAAACGTGCACCTTTTCCACACAAACTCTCTGCTCCCAACT 2478
Db 717 thlsGlyleuIleuAspsasnValhlslPheglInasnserAlaIleleuIleasprule 737
OY 2479 GATCCGAGCGAGAAACCTTACACAGCTTCAGATCTACCCCAAGAGACAGAGATTTTCG 2538
Db 737 uglInasnArglylValAspPheasprleuMetvaltyrproAsnAlaIhlslserleuse 757
OY 2539 CTCGCCGAGTGGCGGAGCAGCATATGAAGTCAAGCTGCTGCACTTCTTACAGGAA 2593
Db 757 rserArgthThrser-----hlslvalAlgllysmetThrhlslPheleuArgln 773
```

```

QY 1135 GCTCCAGCTCGTCCCTCCCGCCCGCCGCTTCATCCCGACAGCAAGTGAAGGCA 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 uValAlaLeuThrIleCysAspPrpAspPrpAlaIleCysArgLeuGluPheUu----- 369
QY 1195 GCGGTAGCTCTGCACAGACTCTCCCGAAGATGCCACCGCTATGTGGTATACAGGA 1254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 -----TyrLysTy 372
QY 1255 GGTCCACCAAGCTGTGATCATATTCATATCATCTTCTATCCCTTCCCAATACAGAGG 1314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 rAlaSerLysArgTrrpAla---ThrlAspAspPheHisSerIle-----ThrlSerPh 389
QY 1315 AGAGAGAGAGCTGTGCTTCTCCCGCCCAATGATGATGACAGCCGGCTTGCCATTTGA 1374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 egluAspPrpIleuPhePheLeuLeuPheProHisAspLysArgAsp-----AsnAlaPh 406
QY 1375 CAATGTCACCGCGTTTAAATCCAGGGCTACGATTGGAGTCCGCTCAGCCCGG 1434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 eglIngluValAlaSerLeuArg-----LeuSerHisgl 417
QY 1435 GGAAGATGATTTAAGTCCCGCCCAATTAAGAAAGATTGCTGTACCCAGCGGTGAATGGA 1494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 yglIn-----LeuArgThrProLysPheLeuAsnLeuLysLysTyrAs 431
QY 1495 GCTTTGGGAGGAGCGGCTCAAGATCTGGTCAATGAGAGACCAAGCTGGTACTT 1554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 pValThrSerIleAsnLys-----IleAsnLysgluThrArgThrIlePhePh 447
QY 1555 CCAGGGCACCAGAGACACCGCGCTGGAGCACACCTCTACGTGTGCTACGTATGAGCGGC 1614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 447 eHisAlaAlaAlaProLysProSerHisArgSerLeuPhe-----SerTyrSerLeuAl 465
QY 1615 CGCGGAGATGTACGGCTCCACAGCGCGCTCTCCCATAGTGGTGCCTCATG----- 1666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 aaSpglu-----SerArgAsnSerAlaTyrCysIleSerCysSerIleLysAsnCy 482
QY 1667 -----AGCCAGAATCTCGACATGTTGCTGACCCACTACAGCAGCAGTGAACAC 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 sThrTrpAlaGluAlaGluMetAspAspLysMetLysThrAlaIleValSerCysLysgl 502
QY 1714 GCGCGCCCTGGCTCAC-----GTTCACAAGCTGAGCGGCCCGCAGCAGCAGCCCTGCA 1767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 yProAlaAlaProHisThrAlaIleValAsnLeuThrArgMetAspSerLysLysTh 522
QY 1768 CAAGCAG-----CCCGGCTCTGGGAGATGATGAGAGCAGCAGCAGCTG 1812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 rgluHisAlaAsnLeuLeuTyrAspLysThrTyrGlnAsnArgValGluGluAlaGlyLe 542
QY 1813 CCGCCCGGATTTATCTCCACAGATCTTCATTTCCACACCGCGCTGGATGTGCGGT 1872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 uPro-----ValIleIleLysgluThrIleLysIleSerAspAspPheAspAlaLeuI 560
QY 1873 CTACGGCATGATGTACAGCCCAAGCGCTTGACAGCCAGGAAGACAC-----CC 1923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 eLysLeuSerIle-----ProLysAspIleTyrAsnArgAspLysHisGlnAlaIlePr 578
QY 1924 CACCGTCTCTTTGTATATGAGAGCGCCCGCAGTGCAGCTGTGAATAACCTCTTCAA-- 1981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 oleuIleValHisValTyrIleLysIleProAsnAspGlnAsnThrLysGluAlaThrGlnI 598
QY 1982 -GGCATCAAGTACTTGGCTCAACACACTGCGCTCCCTGGGTACGCGCTGGTGTG-- 2038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 eglIleIleglu-----GluValValAlaSerAlaSerGlnAlaAlaIleLeuAr 614
QY 2039 -ATTGACGAGAGGGCTCTGTACGAGAGGGCTTCGTTGAAAGGGCGCTGAAAAACA 2097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 614 glleAspPglYArgLysSerLysLysLysTyrArgSerAlaIleTyrLysgl 634
QY 2098 AATGGGCGAGCTGAGATCGAGACCAAGTGCAGGCGCTCAGTGCCTGCGCGAAGTA 2157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 nleuLysIleThrValIleGluValIleLysPglIleLysAlaIleLysValValLeuArgLeuTy 654

```

```

QY 2158 T---GGCTTCATCGACCTGTGACGCCAGTGTGCATCATAGGTGTGCTTACGGGGCTTCC 2214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 rArgHisLeuLeuAspAlaArgArgValAlaValAlaPheGlyTrrpSerTyrGlyLysPheMe 674
QY 2215 CTGCTCATGGGGCTTAATCCACAAGCCC---CAGTGTTCACAGTGGCCATCGCGGCTGC 2271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 tThrLeuSerMetValAsnLysAlaProGluGlnPhePheLysCysAlaValSerValAl 694
QY 2272 CCGCGTACCGCTGTGATGGCCCTACAGCACAGGGTTCACAGTACAGCGGTATACG--GACGT 2328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 aProValThrAsnPheAlaTyrTrpAspAlaThrTyrThrGluArgTyrMetLysPAl 714
QY 2329 CCCTGAGAACACACACAGCAGCGTATGAGCGGGGTCCGTCGCGCTCAGCTGAGAGAGCT 2388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 aPro-----LeuGluSerTyrSerAspValThrLysLysLe 726
QY 2389 GCCCAAT---GAGCCCAACCGCTTGTATCTTCACAGCGCTTCTGACAGAAACGTGCA 2445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 uAspAsnPheLysSerThrArgLeuLeuMetHisGlyLeuLeuAspAspAsnValH 746
QY 2446 CTTTTTCACACAACTCTCTGCTCCCAACTGATCCGACGAGGAAACCTTACAGCT 2505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 746 sPheGlnAsnSerAlaIleLeuIleAspGluLeuGlnAsnArgLysValAspPheAspLe 766
QY 2506 CCAGATCTACCCCAACAGAGACACAGTATTCGCTGCCGAGTCCGGGAGCAGCATATGA 2565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 uMetValTyrProAsnGlnAlaHisSerLeuSerSerArgThrSer-----HisValVa 784
QY 2566 AGTCACGTTGCTCAGCTTCTTACAGGAA 2593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 lGlyLysMetThrHisPheLeuArgLIn 793

```

RESULT 10
T25173
hypothetical protein T23F1.7a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T25173
R:WIKinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25173
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-779 <WIL>
A:Cross-references: EMBL:281129; PIDN:CA803411.1; GSPDB:GN00023; CESP:T23F1.7a
A:Experimental source: clone T23F1
C:Genetics:
A:Gene: CESP:T23F1.7a
A:Map position: 5
A:Intons: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
C:Superfamily: dipeptidyl-peptidase IV

Alignment Scores:
Pred. No.: 3, 32e-20 Length: 779
Score: 467.50 Matches: 205
Percent Similarity: 41.72% Conservative: 120
Best Local Similarity: 26.32% Mismatches: 300
Query Match: 9.75% Indels: 154
DB: 2 Gaps: 43

```

US-09-976-674-4 (1-2617) x T25173 (1-779)
QY 385 GGAAGTGTGAGGAGGAGGAA---AGCGCTGGGGGTCTTGCGCATCAGCTCTAGAGACT 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 GlYArgGluAsnGlyAlaGluIleThrProSerAlaAspArgLysTyrPheAlaMetMet 126
QY 442 CCACAGCAGAGTGGCTCTTCTCTTCCAGGC-----CAGCAAGAGCCTTTCGA 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 AspHis-----AlaProAsnProGlyMetAsnProLIn-AsnGluThrPheH 142
QY 493 CTGTCCGAGCGCGGCGCAAGACGGCTTCATGTGTCCCTATGAACCGCTGGAATCAA 552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 96 AspgluDheglHisSerlleasnAspTyrSerlleSerProaspGluPheIleu 115
QY 470 CAGGCCAACAACAGCTCTTCCACTGTGCGAGCGGCGAAGACGGCTTCATGCTGC 529
Db 116 LeuGluTyrAsnTyrValIysGlnTPrArgHisSerTyrThrIleAspTyrAspIleTyr 135
QY 530 CCAATGAACCGGTGAAATCAAGACCAGCTGCAGGCGCCGAGAGAGACCAAAATC 589
Db 136 AspleuAsnTyrArgGlnLeuIleThrGluGluArgIleProAsnAsnThrGlnTPrVal 155
QY 590 TGCCCTCCGACCCCTGGCTTC---TTCCTCTCAATCAATAACAGCAGCTTGGGTGGC 646
Db 156 ThrTPrSerProValGlnHisLysLeuAlaTyrValITrpsAsnAspIleTyrValLys 175
QY 647 AACATCAGACAGAGCGAGGAGCGGCTGACCTTGCACCAAGGTTTATCCATGTC 706
Db 176 IleGluProAsnLeuProSerTyrArgIleThrTPr-----ThrGlyLysGluAspIle 193
QY 707 CTGGATGACCCCAAGTCTGCGGGGTGGCCACCTTCATACAGAAAG--TTCGAC 763
Db 194 IleTyrAsn-----GlyIleThrAspTPrValTyrGluGluValIlePheSer 209
QY 764 CGCTTCACCTGGTACTGTGGTGGCCCAACAGCTCTCGGAGAGTTACAGAGGCTTCAG 823
Db 210 AlaTyrSerIleLeuTPrTPrSerProAsnGlyThrPhe----- 222
QY 824 ACGCTGCAATCCTGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 877
Db 223 -----LeuAlaTyrIleGlnPheAsnAspThrIleValIleGluTyrSer 239
QY 878 ---GTCCCTCTCTGCTGCTAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934
Db 240 PheTyrSerAspGlnSerLeuGlnTyrProLysThrValArgValProTyrProLysAla 259
QY 935 GCGAGCAAGATCCCAAGATTCGCTGAACCTGCTGATTCGACTGACAGCAGCAGCAGCAG 994
Db 260 GllValAlaIleAsnProThrVal-----LysPhePheValValAsnThrAspSerLeuSer 277
QY 995 AAGATCTCTCGACCCAGGAG 1054
Db 278 SerValThrAsnIleAsnThrSerIleGlnIleThrAlaProIleAsnMetLeuIleGlyAsp 297
QY 1055 GTGGAGTACATCCGCGAGGCGGGGTGAGCCGGGATGCAAAATACGCTGGCGCATGTC 1114
Db 298 ---HisTyrLeuLysAspValThrTPrAlaThrGlnGluArgIleSer----- 312
QY 1115 CTGGACCGGCGCCAGCAGTGGCTCCAGCTGCTCTCTCCCGCGGCTTTCATCCGC 1174
Db 313 -----LeuGlnTPrLeu----- 316
QY 1175 AGCACAAGATGAGAGAGCGGCTAGCCTTCGACAGAGCTTCGCCAGATGTCAG 1234
Db 317 -----ArgArgIleGln 320
QY 1235 CCGTATGTGTCG-----TACGAGAGGTACACCAACGTCGTGATCAATGTT 1279
Db 321 AsnTyrSerValMetAspIleCysAspTyrAspIleSerGlyArgTPr----- 337
QY 1280 CARGACATCTTCATCCCTTCCCAATCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
Db 337 ----- 337
QY 1340 GCCAATGATGACAAGACGGCTTCTGCAATTTGTACAAAGTACACGGCGTTTAAATTC 1399
Db 338 -----AsnGlySerLeuValAlaIleArgGlnHisIleGluMetSerThrTPr 351
QY 1400 CAGGGCTACGATTTGGAGTGAAGCCCTTCAGCCCGGAGAGATGA----- 1444
Db 352 -----GlyTPrValGlyArgPheArgProSerGlnProHisPheThrLeuAspGly 368
QY 1445 -----TTTAAAG 1450
Db 369 AsnSerPheTyrLysIleIleSerAsnGluGluGlnTyrTArgHisIleCysTyrPheGln 388

QY 1451 TGCCCAATTAAAGAGAGATTGCTGTACACACCGGTGAATGGAGGTTTGGCAGGAC 1510
Db 389 IleAspLysLysAspCysThrPheIleThrLysGlnTPrGluValIleGlyIleGlu 408
QY 1511 GCGTCCAGATCTGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGCGCAAGAGAC 1570
Db 409 AlaLeu----- 1570
QY 1571 ACGCGCTGGAGCACCAACCTCTAGTGGTCACTATGAGCGGCGCGC----- 1618
Db 414 -----TyrLeuTyrTyrIleSerAsnGluTyrLysGlyMetProGlyGly 428
QY 1619 ---GAGATCGTACCGCTCACACACCGCGCGCTTCCCATAGCTCTGTCAGACCAAAAC 1675
Db 429 ArgAsnLeuTyrLysIleGlnLeuIleAspTyrThrLysValIThrCysLeuSerCysGlu 448
QY 1676 TTGCAATGTTTCGTACAGCCACTACAGAGAGCGTGAGACCGCGCGCTGTCACGTCAC 1735
Db 449 LeuAsnProGluArgCysGlnTyrTyrSerValSerPheSerLysGluAlaLysTyrTyr 468
QY 1736 AAGCTG-----ACGCGCGC----- 1750
Db 469 GlnLeuArgCysSerGlyProGlyLeuProLeuTyrThrLeuHisSerSerValAsnAsp 488
QY 1751 -----GACGACAGCCCTTCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
Db 489 LysGlyLeuArgValIleGluAspAsnSerAlaLeuAspLys----- 502
QY 1790 ACGATGATGAG 1849
Db 503 ---MetLeuGlnAsnValGln-----MetProSerLysLysLeuAspPhe 516
QY 1850 CACACGGCTCGGATGTGGCGTC---TACGGCATGATCTACAGAGAGAGAGAGAGAGAGAG 1906
Db 517 IleIleLeuAsnGlnTPrLysPheTyrTyrGlnMetIleLeuProPheHis---PheAsp 535
QY 1907 CCAGGAGAAAGACACCCAGCGCTCTTGTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1966
Db 536 LysSerLysLysTyrProLeuLeuLeuAspValTyrAlaGlyProCysSerGlnLysAla 555
QY 1967 AATATCTCTCAAAAGCATCAAGTACTTGGCGCTAAC-----ACATGGCGCTCC 2017
Db 556 AspThrValPhe-----ArgLeuAsnTPrAlaThrTyrLeuAlaIleSer 569
QY 2018 CTGGGCTACCGCGCTGCTGTC---ATTGACGCGAGGCGCTCTTCAGACGAGCTTCGG 2074
Db 570 ThrGluAsnIleValAlaIleSerPheAspPheLysArgGlySerGlyTyrGlnGlyAspLys 589
QY 2075 TTGCAAGGCGCCCTGAAGAAACCAATGAGCCAGGTGAGATCGAGAGAGAGAGAGAGAGAG 2134
Db 590 IleMetHisAlaIleAsnArgArgLeuGlyThrPheGluValGluAspGlnIleGluAla 609
QY 2135 CTG---CAGTTCGTGGCGGAGAGAGTATGCTTCATTCAGCTGAGCCGATTCATCAT 2191
Db 610 AlaArgGlnPhe-----SerLysMetGlyPheValAspAsnLysArgIleAlaIleTPr 627
QY 2192 GCGTGTCTACAGGCGGCTCTCTGCTCATAGGGGTAAATCCAGAGCCGAGAGTTC 2251
Db 628 GlyTPrSerTyrGlyGlyTyrValThrSerMetValIleuGlySerGlySerGlyValPhe 647
QY 2252 AAGGTGCGCATCGCGGGGTGGCCCGGCTGATGAGTGGCTTCAGACAGAGGATCACT 2311
Db 648 LysCysGlyIleAlaValAlaIleProValSerArgTPrGluTyrGlyAspSerValTyrThr 667
QY 2312 GAGCGCTACATGAGAGCTC-----CTGAAGAACACACAGAGCGGCTATGAGCGGCTTC 2365
Db 668 GlnTyrGlyMetGlyLeuLeuProThrProGluAspAsnLysAsnIleTyrArgAsnSerThr 687
QY 2366 GTGGCGCTGACAGGAG 2425
Db 688 ValMetSerArgAlaGluAsnPhe-----LysGlnValGluTyrLeuLeuIleHisGly 705

Db 602 pheansnglyasnAlaphargTyrseValserValrghnIsleuglyglutrlpGlusertYtr 621

Oy 2120 GACCAAGGTGGAGGGCCCTCACTTCGTGGCCAGAGAAGTATGGCTTCATGACCTAGACCCGA 2179

Db 622 AspglnlglYnlnlaaglyIsphetrpAlaasp--LeuProPheValAspIslaHis 640

Oy 2180 GTTGCATCCATAGGTGGTGTCTACAGGGGCTTCCTCGTCATAGGGGCTATACACAAg 2239

Db 641 ValglYlIetrpGlytrpserTyrclYglYtrleuthrleuysThrleu---GluThr 659

Oy 2240 CCCAAGTGTTCOAAGTGGCCATGCGGGGTGCCCGGTCCACCGTCTGGATGGCCCTAGAC 2299

Db 660 GluAspValPheSerTyrclYmetAlaValAlaProValAlaPsrTyrArgLeuTyrAsp 679

Oy 2300 ACAGGGTCACTAGAGCGGTACATGAGCTCCCTGAGAACACCAACGACGCGTATAGGGCG 2359

Db 680 SerValYtrThrIgluAlgtYrMetAspLeuProGlnTyrAsnIysglYlYrIylAsn 699

Oy 2360 GGTTCGTGGGCCCTTCACAGTGGAGAGAGCTGCCCAATGAGGCCCAACCGCTTCATATCCTC 2419

Db 700 SerGlnIle-----HisAspTyrclulysPheIysGlnleuIysArgPhePheAla 717

Oy 2420 CACGGCTTCCTGGAGCAAGAGTGCATCTTTTCCACACAAATCTTCGTCTCCCACTG 2479

Db 718 HisglYtrIglYAspAspAsnValHisPheGlnHisSerMetHisIleuMetAspIylleu 737

Oy 2480 ATCCGACGAGGG---AAACCTTACGACTCCAGACTCCCAACGACAGAGACAGATTT 2536

Db 738 AsnleuAlaAsnGysTyrIAsnTyrIAspMetAlaValAlaPheProAspSerAlaHisSerIle 757

Oy 2537 CGCTGCCCCCGATCGGGCGAGCACTATGAGTACAGTTCGTCTTCATCCAGAAATAC 2596

Db 758 -----SerTyrHisAsnAlaSerLeuSerIleTyrHisArgLeuSerclutrp 773

Oy 2597 CMC 2599

Db 774 Ile 774

RESULT 6

CDHU26

N.Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)

C.Species: Homo sapiens (man)

C.Date: 31-Dec-1993 #sequence-revision 23-Aug-1996 #text-change 18-Jun-1999

C.Accession: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520

R.Biochem. Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.

R.Biochim. Biophys. Acta 1131, 333-336, 1992

A.Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a serine protease

A.Reference number: S24313; MUID:92329551; PMID:1352704

A.Accession: S24313

A.Molecule type: mRNA

A.Residues: 1-6, '1', 8-766 <MIS>

A.Cross-references: EMBL:X60708; NID:935335; PIDN:CAA43118.1; PID:935336

R.Darmoul, D.; Lacesa, M.; Barilaault, L.; Marguet, D.; Sapin, C.; Trotot, P.; Barbat, A.

J. Biol. Chem. 267, 4824-4833, 1992

A.Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines

IV mRNA levels during cell differentiation.

A.Reference number: A42408; MUID:92165847; PMID:1347043

A.Accession: B42408

A.Molecule type: mRNA

A.Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DARI>

A.Cross-references: GB:M60536; NID:9181569

A.Note: sequence extracted from NCBI backbone (NCBI:83986; NCBI:83988); this sequence is identical to the one in the GenBank database.

R.Corvel, J.P.; Ferrero, A.; Chambrand, L.; Rigal, A.; Bonicel, J.; Maroux, S.

Gastroenterology 101, 618-623, 1991

A.Title: Expression of sucrase-isomaltase and dipeptidyl peptidase IV in human small intestine

A.Reference number: A61136; MUID:91317403; PMID:1677636

```

A:Accession: B61136
A:Molecule type: protein
A:Residues: 1-15, 'X', 17-22 <GOK>
R:Boehm, S.K.; Gun Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.
Biochem. J. 311, 835-843, 1995
A:Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from
A:Reference number: S59510; MUID:96067599; PMID:7487939
A:Accession: S59510
A:Molecule type: DNA
A:Residues: 1-31 <BOE>
A:Cross-references: GB:S79876; NID:g1195574; PIDN:AA35614.1; PID:g1195575
R:Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg
J. Immunol. 149, 481-486, 1992
A:Title: Cloning and functional expression of the T cell activation antigen CD26.
A:Reference number: I56154; MUID:92325476; PMID:1352530
A:Accession: I56154
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-436, 'S', '438-766 <TAN>
A:Cross-references: GB:M4777; NID:g180082; PIDN:AA51943.1; PID:g180083
R:Abbot, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.
Immunogenetics 40, 331-338, 1994
A:Title: Genomic organization, exact localization, and tissue expression of the human
A:Reference number: S59857; MUID:95012454; PMID:7927537
A:Accession: S59857
A:Molecule type: DNA
A:Residues: 1-436, 'S', '438-766 <ABB>
A:Cross-references: EMBL:U13734
C:Genetics:
A:Gene: GDB:DDP4
A:Cross-references: GDB:125239; OMIM:102720
A:Map position: 2q24.3-2q24.3
A:Introns: 2/3; 32/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 35
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidyl-peptidase; glycoprotein; homodimer; proteinase; transme
F:1-6/Domain: intracellular #status predicted <INT>
F:7-28/Domain: transmembrane #status predicted <TM>
F:129-766/Domain: extracellular #status predicted <EXT>
F:85,92,150,219,229,281,321,520,685/Binding site: carbohydrate (Asn) (covalent) #statu
F:630,708,740/Active site: Ser, Asp, His #status predicted

Alignment Scores:
Pred. No.: 2.08e-21 Length: 766
Score: 487.50 Matches: 201
Percent Similarity: 36.27% Conservative: 124
Best Local Similarity: 22.43% Mismatches: 330
Query Match: 10.17% Indels: 241
DB: 1 Gaps: 35

US-09-976-674-4 (1-2617) x CDH026 (1-766)
QY 149 CTCATTGTCACACAGCGCCGCCACGACTTCAGTTGTGACAGAACGATGAGTCTGGG 208
:::|||||
Db 26 ValLeuValnSLyGly-----ThrAspSplathr 36
QY 209 CCCCACTCCACCGGCTCTACTACTCTGGGAATGCCATATGGACGCCGAGAACCTCC 268
|||:::||||
Db 37 AlaaspeSearGlySfrHtyrThrLeu----- 45
QY 269 CTCCTCTCTGAGATTCCCAAGAGAGTCCGGAAGAGGCTCTGCTCTCTCTGGAAG 328
:::::|||||
Db 46 -----ThrAspTyLeuLysasnthrTyArGleuLysleuTySerLeuAlgrlp--- 62
QY 329 CAGATGCTGGATCATTTCCAGGCCACGCCCCACCCATGGGGTCTACTCTCGGAGAGAG 388
:::|
Db 63 -----IleSerAspHisGluTyLeuTyfLysGlnGluAsn 75
QY 389 CTGTCTG-----AGGACGCGGAACGCCTG 412
:::||||
Db 76 ILeuValAlphasAlaGluTyrlGlyLysnSerSerValPheLeuGluLysnSerThrPhe 95
QY 413 GGGGCTTCAGGATCACCTCTACGACTTCCAC---AGCGAGATGGGCTCTTCCTTTC 469
|||||||:::|||||:::|||||

```

Db 645 -----LysGlyLysPheLeuLeuIleHisGlyThrAlaAspAsnValHisPhe 661
 QY 2450 TTCACACAAACTCTCGTGTCTCCCACTGATCCGAGAGGGAACCTTACCAGCTCCAG 2509
 Db 662 GlnAsnSerMetGluPheSerGluAlaLeuIleGlnAsnLysGlnPheAspPheMet 681
 QY 2510 ATCTACCCCAAGAGAGACACAGTATT 2536
 Db 682 AlaTyrProAspLysAsnHisSerIle 690
 RESULT 5
 T41703
 dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T41703
 R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: 222011
 A:Accession: T41703
 A:Status: preliminary; translated from GB/EMBL/DDBT
 A:Molecule type: DNA
 A:Residues: 1793 <MOR>
 A:Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
 A:Experimental source: strain 972h-; cosmid c2E11 -chimeric
 C:Genetics:
 A:Gene: SPAC2E11.08
 A:Map position: 1
 C:Superfamily: dipeptidyl-peptidase IV
 Alignment Scores:
 Pred. No.: 1.16e-22 Length: 793
 Score: 508.50 Matches: 178
 Percent Similarity: 39.66% Conservative: 100
 Best Local Similarity: 25.39% Mismatches: 274
 Query Match: 10.60% Indels: 149
 DB: 2 Gaps: 25
 US-09-976-674-4 (1-2617) x T41703 (1-793)
 QY 614 TCCTTCATCAATTAACAGCAGCTGTGGTGCCCAATCGAGAGCGAGAGCGGCG 673
 Db 184 SerPheValIyrAsnAsnAspIeutyValArgLys---AsnAspGlyAsnValGlnArg 202
 QY 674 CTGACCTTCGCCACCAAGGTTTATCCAAATCTCTGATATGCCCAAGTCTGGGGGTG 733
 Db 203 LeuThrTyr-----AspGlyThrValAspValPheasn-----GlyLeu 215
 QY 734 GCCACCTTCGTCATACAGAAAG---TTCGACCGCTTCACACTGGGTACTGGTGGCC 790
 Db 216 ThrAspTrpIleTyrGluGluGlnValLeuSerSerProSerTrpIleTrpTrpSerPro 235
 QY 791 ACAAGCTCTGGGAAGGTTCAAGAGGCGCTCAAGAGCTCGCAATCTGTATGAGGAAGTC 850
 Db 236 AspSerAsp-----LysIleAlaPheLeuLeu 245
 QY 851 GATGAGCCGAGGTGAGGTGATTCAGTCCCG-----TCTCTGGC 892
 Db 246 AsnGlnSerGluIleProThrTyrHisTyrProLeuTyrThrAlaGluLeuAspProSer 265
 QY 893 CTAGAAGAA-----AGGAAGAGCGACTGCTGATCCAGACTGACAGCCAGGCGAAGATGTC 1003
 Db 266 LeuProGluIleAspTyrAsnLysAspMetAlaIleLysTyrProLysProGluLysPro 285
 QY 944 AATCCCAAGATTGCTTGAATCTGGCTGAGTTCAGACTGACAGCCAGGCGAAGATGTC 1003
 Db 286 AsnProSerValSerLeuPheValAlaAspLeuAsnSerAsnAlaSerSerAsnPheSer 305
 QY 1004 TCACACCAAGAGAAGAGCTGGTGCACCGCTTCAGCTCGCTTCGCCGAAGGTGGAATAC 1063
 Db 306 LeuTrpHisAsnGluProValLeuAlaGluPro-----Val 316
 QY 1064 ATCGCCAGGCGCGGTGGACCCGGGATGGCAAAATACGCTGGGCCATGTTCCTGGACCG 1123

Db 317 ValGlnAsnValLeuTrpValAsnThrSerSer---ValLeuValGlnPheThrAsnArg 335
 QY 1124 CCCCAGCAGTGGCTCCACAGTCTCTCTCTCCCCCGGCGCTTCATCCAGACAGAG 1183
 Db 336 AsnSerThrCysIleThrAlaArgLeuLeuAspThrGluLysSerIleHisThrVal 355
 QY 1184 AATGAGGAG-----CAGGGCGTACGCTGGCCAGAGCGTGTCCC 1222
 Db 356 LysThrGluCysLeuGluGluGlyTrrTyrGluValGlnGlnSerAlaLysMetPhePro 375
 QY 1223 AGGAATGTCCAGCCGATGTGGTGTACAGAGGCTCAACAGCTGTGATCATGTTCAT 1282
 Db 376 Leuasn-----AsnSerLeuValTrpGluAsnTrpSer 386
 QY 1283 GACATCTTCTATCCCTTCCCAATTCAGAGGAGAGAGCAGCTGTCTTCTCCGGCC 1342
 Db 387 AspGlyTyrPheAspIleLeuAlaLeuAspAspTyrAsnHisLeuAlaPheIle----- 404
 QY 1343 AATGAATGCAAGACCGGCTTCCCATTTGTACAAAGTCACCGCGTTTAAATCCAG 1402
 Db 404 ----- 404
 QY 1403 GGTACGATTGGAGTGAACCTTCAGCCCGGGAAGATGAATTTAAGTCCCATTAAG 1462
 Db 405 -----ProPheAsnGly-----Ser 409
 QY 1463 GAAGAGATTGCTCTGACACGAGGCTGAATGGAGGTTTGGCAGGACGCTCCAGATC 1522
 Db 410 SerProIleTyrLeuThrSerGlyAlaTrpAspValThr-----AspGlyProIle 426
 QY 1523 TGGTCAATGAGGAGACCAACAGCTGTGACTTCCAGGGCACCAAGAGACGCGCTGGAG 1582
 Db 427 HisIleAspGlyAspPheGlyAsnValTyrPheLeuAlaTrpLeuLysAspSerThrGlu 446
 QY 1583 CACCACCTTCAGTGTGACGCTATGAGGCGCGCGAGATCTGACGCTTACACGACGCGCC 1642
 Db 447 ArgHisLeuTyrTyrValSerLeuAspThrLeu---GluIleTyrGlyIleThrAspAsn 465
 QY 1643 GCGTTCTCC-----CATAGCTGCTCCATGAGCCAGAACTTCACATGCTGTCAGC 1693
 Db 466 GlnGluAspGluGlnGlyTyrTyrSerThrSerPheSerProPheGlyAspPheThrValLeu 485
 QY 1694 CACTACAGCAGCGTGACAGCGCGCGCTGCGTGCACGCTTACAGCTGAGCGGCGCGAGC 1753
 Db 486 AsnTyr-----HisGlyProAsp 491
 QY 1754 GACAGACCCCTGCACAGACAGCCCGCTGCGCTGAGCATGATGAGAGCGACGCTGC 1813
 Db 492 ValPro-----TrpGlnIleuLeuArgSerThrLysAsp--- 502
 QY 1814 CCCCCGATTAT----- 1825
 Db 503 ---LysAspTyrCysLeuSerLeuGluThrAsnSerArgLeuLysGlnLeuSerSer 521
 QY 1826 -----GTCTCCAGAGATCTTCATTTCCACAGCGCGCTGCGGTGCGGCTTACGCGC 1879
 Db 522 IleThrLeuProSerValGluTyrGlyLysLeuThrPheAsnAspThrThrPheAsnPhe 541
 QY 1880 ATGATCTTACAAAGCCCGACGCTTCGACAGGGAAGAGCCCGACGCGCTCTTGTGA 1939
 Db 542 MetGluArgTrpProAlaGlnPheAspValAlaAsnLysLysTyrTrpValLeuPhePheAla 561
 QY 1940 TATGAGGCGCCCGCAGGTGACGCTGTGAATTAATCTTCAAAAGCATCAAGTACTTGGC 1999
 Db 562 TyrGluGlyProGlySerGlnGlnValAlaLysLeuPheArgValAspPheGlnAlaTyr 581
 QY 2000 CTCAACACACTGGCGCTCCGCTGAGCGGCTGAGCGGTGTGTATGACGCGAGGGCTCGT 2059
 Db 582 LeuAlaSerHisProAspPheGluPheIleValValThrLeuAspGlyTyrGlyThrGly 601
 QY 2060 CAGCGAGGCGCTGCTTCGAAGGGCGCTGAAAAAACCAATGGGCGAGGTGAGATCGAG 2119


```
Db 382 LeuGluGly-----ValAspGluAlaArgLysValAlaIlePheSerAlaSer 397
1565 AAGGACAGCGCGGTGGAGCACCACCTCTACGTGAGCATGAGCGCGCGAGATC 1624
Db 398 IleAspThrProIleGluArgLeuTyrGluValSerTyrAlaLysProGlyLysPro 417
1625 GTAGCGCTCAGCAGCGCGGCTTCCCATAGCTGCTCCATGAGCCAGAACTTGCAGATG 1684
Db 418 LysAlaLeuThrSerIleGlyGlyTyrTrpTrpAlaIleValAlaLysAspGlyAla 437
1685 TTGTCAGCCACACACAGCAGCGGTGAGCGCGCGCTGACAGCTTTCAGAGCTGAGC 1744
Db 438 PheIleGlyThrTyrSerProLysTrpProSerGlnThrAlaLeuTyrSerAlaAsp 457
1745 GGGCGC-----GAGCAGACCGCGCTGACAGCAGCGCGCGCTGAGGCT 1789
Db 458 GlyLysArgValArgTrpIleGluLysLysLysLeuAlaGluLysProTyrTrpPro 477
1790 AGCATGATGGAGCAGCCAGCTGCGCGCGGATTAATGTTCCAGAGATCTTCATTTC 1849
Db 478 -----TyrAlaIleAsnLeuPro-----GlnProGluPheGlySerLeu 490
1850 CACAGCGCTCGATGCGGCTCTACGGCATATCTACAGCCCAAGCGCTTGACGCCA 1909
Db 491 LysAlaIleAspGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 510
1910 GGGAGAGAGCGCGCGCGCGCTTGTATGAGAGCGCGCGCGCGCGCGCGCGCGCGAAT 1969
Db 511 AlaLysLysTyrProAlaIleValSerValTyrGlyLysProIleAlaGlnArgValMet 530
1970 AACTCTTCAAGGACATCAAGTACTTGGGCTCAACACACTGCGCTCCCTGGGCTACGCC 2029
Db 531 LysAsnTrpIleSerProSerGlu-----ArgThrTyrLeuGluAlaGlyTyrVal 547
2030 GTGCTTGTGATTGAGCGAGCGGCTCTGACCGAGGCTCGGTTGCAAGGCGCTG 2089
Db 548 IlePheLysLeuAspAsnArgGlySerGlyAsnArgSerAlaLysPheMetArgAlaLeu 567
2090 AAAAACCAATGGCGAGGATGCGAGCAGGATGAGCAGGATGAGCGCGCTGCGGCGC 2149
Db 568 AspArgLysLeuGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 587
2150 GAGAGTATGCTTCATGCACTGAGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 2209
Db 588 SerGln---ProTyrValAspAlaAspLysLeuGlyValMetGlyTrpSerTyrGlyGly 606
2210 TTGCTCTCGCTTCATGCGGCTATGCAAGCGGATGCGGATGCGGATGCGGATGCGG 2269
Db 607 PheMetAlaLeuMetLeuLeuThrAlaGlnAsnThrProPheLysAlaGlyAlaIleGly 626
2270 GCGCGGCTCAGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 2329
Db 627 AlaProThrThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 646
2330 CCGTGAACAACAGCAGGATGAGGCGGATGAGGCGGATGAGGCGGATGAGGCGGATGAG 2389
Db 647 ProAspGluAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 666
2390 CCGAATGAGCGGATGAGGCGGATGAGGCGGATGAGGCGGATGAGGCGGATGAGGCG 2449
Db 667 -----AlaProGlySerLeuLeuLeuIleMetAlaAspAsnValIlePhe 684
2450 TTTCACACAACTTCTCGTCTCCAACTGATGCGGATGCGGATGCGGATGCGGATGCGG 2509
Db 685 GluAsnSerThrArgLeuMetAlaIleValGlnArgLysAlaIleLeuPheGluMetAla 704
2510 ATCTACCCCAAGCAGAGCAGCAGT 2533
Db 705 MetTyrProGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 712
RESULT 3
JCS142
```

```
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C:Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C:Accession: JCS142
R:Kashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl-peptidase IV from Xanthomonas maltophilia: Sequencing and express
A:Reference number: JCS142; MUID:97164011; PMID:9010758
A:Accession: JCS142
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-References: DDBJ:D83263; NID:g1753196; PIDN:BA11872.1; PID:g1753197
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted <TM>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted
```

```
Alignment Scores:
pred. No.: 1.03e-38 Length: 741
Score: 776.00 Matches: 238
Percent Similarity: 44.86% Conservative: 133
Best Local Similarity: 28.78% Mismatches: 295
Query Match: 16.18% Indels: 161
DB: 2 Gaps: 32
```

US-09-976-674-4 (1-2617) x JCS142 (1-741)

```
QY 202 GTGTCGGCCCGCCATCCACCGCGCTCTACTGAGGATGCAATATGAGCGCGAGAGA 261
Db 28 IleThrGlyProLeuPro-----LeuSerGlyProThrLeuMetLysProLysVal 44
QY 262 CTGCTCTCTACTGATGATGCC-----AACAGGTCGCGAAGAGCGCTCT 309
Db 45 AlaProAsp-GlySerArgValThrPheLeuArgLysAspSerAspArgGlnLeu 64
QY 310 GCGTCGCTGCTCTGCG-----AACGACATGCTGATCATTTCCA 348
Db 64 ValPheLysPheTyrAspIleGlySerGlyGlnThrArgLeuLeuValAspSerLysVal 84
349 GCGCAGCGCCCGCCATGAGGCGCTACTCTCGGAGAGAGAGCTGAGGAGCGGAACG 408
Db 84 ValLeuProGlyThrGluThrLeuSerAspGluGluLysAlaArgArgLysGlnArg 104
QY 409 CTTGGGGGCTTC---GCGATCAGCTCTGACACTTCACAGCGAGAGTGGCTTCT 465
Db 104 GlnAlaAlaMetThrGlyLysValAspTyrGlnTrpSerProAspAlaGlnArgLeu 124
QY 466 CTTTCAGCGCGAGCAAGC-----CTCTTCAGCTCGCGGAGGCGGAGCAAGCGCTT 519
Db 124 ValPheLeuGlyGlyLysLeuTyrLeuTyrAspLeuLysGlnGlyLys----- 141
QY 520 CATGTCCTCCCTATGAACCGCTGGAATCAAGACCGAGCTGTCAGGCGCGGATGGA 579
Db 142 -----AlaAlaValArgLysLeu-----ThrHisGlyGlnGlyPheAlaThrAs 156
QY 580 CCGCAAAATGTGCGCTGCGGACCGCTTCTCTTCATGATAAATACGCGACTGTG 639
Db 156 ValLysLysSerProLysGly---GlyPheValSerPheIleArgGlyArgAsnLeuTr 175
640 GGTGGCAACATGAGAGCAGGAGCGGCGGCTGAGCTTGTGCGACCAAGCTTATC 699
Db 175 PValIleAspLeuAlaSerGlyArgGlnMetGlnLeuThr-----AlaAspLysThr 193
QY 700 CAATGTCCTGATGACCCCAAGTCTGCGGCTGCGCACCTTCTGATACAGGAAGATT 759
Db 193 rThrIle-----GlyAsnGlyLeuAlaGluPheValAlaAspLysGluMet 208
QY 760 CGACCGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 819
```

```

Db      781 HisGlyTyrPserTyrGlyGlyTyrMetAlaLeuGlnMetIleAlaLysHisProAsnIle 800
OY      2249 TTCAGAGTGGCCATCGGGGGGGCCCGGTCAACGCTGTGATGGCTCGACGACAGAGGCTAC 2308
Db      801 TTTATGTAAlaAlaIleAlaGlyGlyAlaValSerAspTrpArgLeuTyrAspThrAlaTyr 820
OY      2309 ACTGAGGCTACATGATGAGTCCCTGAGACACACGAGCAGGCTATGAGCGGGTTCGCTG 2368
Db      821 ThrGlnArgTyrMetGlyTyrPro---LeuGlnGlnHisValTyrGlyAlaSerSerIle 839
OY      2369 GCCCTGCACCTGAGAAAGCTGGCCCAATGAGCCCAACGCTGTCTTATCTCCACGCGCTTC 2428
Db      840 ThrGlyLeuValGlnTyrLeuProAspGluProAsnArgLeuMetLeuValHisGlyLeu 859
OY      2429 CTGAGCAAAACCTGACCTTTTCCACAAACCTCTCGTCTCCCAATGATCCGACGCA 2488
Db      860 MetAspGlnAsnValHisPheAlaHisLeuThrHisLeuValAspGlnCysIleLysLys 879
OY      2489 GGGAAACCTTACACAGCTCCAGATCTACCCACAGAGACAGACAGATATTCGCTGCCCGAG 2548
Db      880 GlyLysTrpHisGlnLeuValIlePheProAsnGlnArgHisGlyValAlaArgAsnAsp 899
OY      2549 TCGGCGCAGCATATGAACTCAGCTGCTGCACCTTTCTACAGAA 2593
Db      900 AlaSerIleTyrLeuAspAlaArgMetCysTyrPheAlaGlnGln 914

```

RESULT 2

```

AB7516
dipeptidyl peptidase IV [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: AB7516
R:Netman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.;
  B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolton
  n, J.; Ermolenko, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
  Proc. Natl. Acad. Sci. U.S.A. 98, 4116-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: AB7516
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-738 <STO>
A:Cross-references: GB:AE005673; NID:g13423647; PIDN:AAK24125.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2154

```

```

Alignment Scores:
Pred. No.: 1,74e-40 Length: 738
Score: 805.50 Matches: 227
Percent Similarity: 43.78% Conservative: 118
Best Local Similarity: 28.81% Mismatches: 308
Query Match: 16.80% Indels: 135
                Gaps: 25

```

US-09-976-674-4 (1-2617) x AB7516 (1-738)

```

OY      209 CCCACCTCCACGCGCTCTACTACCTG---GSAATGCCATATGGACGCCAGACAACTCC 265
Db      47 ProAspGlyLysArgValThrTyrLeuLysGlyLysProGlnAlaAlaAsnIleGlnAsp 66
OY      266 CTCCTCTACTCTGAGATT-----CCCAAGAAAGTCCGGAAGAGGCTCTGCTG 313
Db      67 LeuTrpAlaAlaAspValLysGlyGlnProTyrArgLeuIleAspSerAlaAlaLeu 86
OY      314 CTCCTCTCTGAGACGATGCTGATCTTCACGAGCCAGCCCAACCATGGGGTCTAC 373
Db      87 SerSerGlyAspLysGlnLeu----- 93
OY      374 TCTCGGAGAGAGAGCTGCTGAGAGAGACGGAACGCTGGGGCTTCGGCATCACCTCC 433
Db      94 SerGlnAlaGlnLysAlaArgValArgGlnAlaArgValSerAlaArgGlyIleValGln 113

```

```

OY      434 TACAGTTCACACAGAGAGTGGCTCTCTCTTCCAGGCGCACAAACGCTCTCCAC 493
Db      114 TyrSerTrpAspArgGlnGlyArgPheIleLeuValProLeuAspGlyAspLeuTyr--- 132
OY      494 TGTGCGACGGCGGCAAGAACGCTTCATGGTGTGCCCTATGAACCGCTGGAATCAAG 553
Db      133 ---LeuAspAlaValAlaAspGly-----LysIleThrArgLeu 144
OY      554 ACCGAGTGTCAAGGGCCCCGATGACCCCAAAATCTGCCTCGCCAGACCTGCTTCTTC 613
Db      145 ThrGlnTrpProGlyAspGlnValAlaSpAlaLysValSerProLysGly---GlyTyrVal 163
OY      614 TCTTCATCATATAACAGCAGCCTGTGGTGGTGGCAATCGACAGACGAGGAGGCGCG 673
Db      164 SerTyrValArgAspGlnAsnLeuTyrIleLysProValAlaGlyGlyAlaGlnThrAla 183
OY      674 CTGACCTTCTGCCACCAAGCTTATGCCAATGCTCGATGACCCCAAGTCTGCGGTGTG 733
Db      184 LeuThrTrpAspGlyLys-----AspAlaLeuSerPheGlyVal 196
OY      734 GCCACCTTCGTATACAGAAAGAGTTCGACCCGCTTCACTGGGTACTGTGTGCCCCACA 793
Db      197 AlaGlnPheIleValGlnGlnGlnLeuAspArgPheThrGlyTyrTrpTrpSerProAsp 216
OY      794 GCTCTCTGGGAAGGTTCAGAGGGCTCAAGACGCTGCGAAATCTGTATGAGGAATGCAT 853
Db      217 GluSer-----ArgIleValTyrThrArgValAsp 226
OY      854 GAGTCCGAGGTGAGGTATTCACGTCCCTCTCTCGCTAGCAAGAAAGAGAGCGAC 913
Db      227 GluSerGlyValAspIleValProArgAlaAspIleGlyProGlyGlyAlaThrValVal 246
OY      914 TCGATGGATACCCAGCAGCAGCAGCAGCAAGAACCCCAAGTGGCTGAAACCTGCTGAG 973
Db      247 AsnGlnTrpArgProAlaGlnGlyArgProAsnAlaValAlaAspLeuPheValAlaArgAsp 266
OY      974 TTCAGACTGACACCGCAGGCGCAGAGTGTCTGCAGCCAG-----GAGAAGAGACTG 1024
Db      267 Leu-----AlaSerGlyLysValThrAlaLeuAspLeuGlyAlaAlaAsnLysAspIle 283
OY      1025 GTGCAGGCTTTCACGCTCGCTTCCCGAAGGTGAGTATCGCCAGCGCGGGTGGACC 1084
Db      284 -----TyrValAlaArgValAlaTrpSer 291
OY      1085 CGGATGCAATATACGCTGGGCCATGCTCTGACCGCGCCAGCAGTGGCTCCACTCTC 1144
Db      292 AlaAspGlyLysThrValTyrValGlnArgLeuSerArgAspGlnLysThrLeuAspLeu 311
OY      1145 GTCTCTCTCCCGCGGCTGTTCATCCGAGCAGACAGATGAGAGCAGCGCTAGCC 1204
Db      312 LeuAlaPheAspAlaAla-----ThrGly 319
OY      1205 TGTGCGAGAGCTGTCCCGAAGATGCCAGCCGTATGTGTACGAGAGGTCCACCAC 1264
Db      320 AlaGlyLysThrIleLeuThrAspThrAspProHisPheIle-----GluValSerAsn 337
OY      1265 GTCTGATCATATGTATCATCTTCTATCCCTTCCCAATCAAGAGAGAGAGAGAG 1324
Db      338 AspPhe-----ArgProLeuThrAspGly----- 345
OY      1325 CTCCTGCTTCTCCGCGCAATGATGACAGACGCGCTTCTGCAATTTGTAAGTAAGTCC 1384
Db      346 ---ThrPheLeuTrpGlySerGln---LysAspGlyAsnGlnHisLeuTyrArgTyrAla 363
OY      1385 GCCGTTTAAATCCAGGCTACGATTGAGTGAAGCCCTTTCAGCCCGGGGAAGATGAA 1444
Db      364 Ala-----AspGly----- 366
OY      1445 TTTAAGTCCCATTAAGAGAAAGATTGCTCTACCAACGCGGTAAATGGAGTTTGGCG 1504
Db      367 -----LysLeuIleAlaGlnIleThrLysGlyAspTrpProValIleGly 381
OY      1505 AGCAGCGGTCCAAAGATCTGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGGCACC 1564

```


OY	320	TCCTGGAGACAGATGCTGGATCATTTCCAGACGCCGCCACCACGAGGGTCTACTCTGG	379
Db	122	SerGlyTyrAsnValaLapSerTyrIleArgMetSerCysArgLysThrProProSerAla	141
OY	380	GAGAGAGAGCTGCTGAGGAGCGGAACGCGCTGGGGGTCTC---GGATTCACCTTCAC	436
Db	142	GluPheThrLeuGlnCysGluValGlnArgSerGlnValValThrGlyIleSerAspTyr	161
OY	437	GACTTCACACAGCGAGGTGGCTCTTCCTTCACAGGCCAGCACAGCCTCTTCAC--	493
Db	162	GluIle---ArgAsnGlyLysMetIleLeuMetalGlyLysPnlLeuPheArgTyr	179
OY	493	-----	493
Db	180	AsnProLeuAsnGluAlaLeuAlaAlaIleProIleAlaValProAspAspGlnSerSer	199
OY	494	-----TGTCGGCAC	502
Db	200	ThrGluProMetAspIleSerGlnGlySerIleThrSerGlyThrLysGlySerSerAsn	219
OY	503	GCGCGCAAGAACGGCTTCATGCTGTCCCTCATGAACCCCTGGAAATCAG-----	553
Db	220	GluAlaProGlnSerSerThrValProProValThrArgIleProIleLysLysProThr	239
OY	554	ACCCAGTGCCTGACGGCGCGGAGTGAGCCC-----AA	586
Db	240	ThrSerThrGlnLysProAlaThrAlaProProThrAsnAsnPheValSerSerAlaLys	259
OY	587	ATGTGCCCTGCGCACCTGCTCTTCCTTCATCATATACAGCGACCTGTGGGTGCC	646
Db	260	ValCysProAlaAspSerSerIleuAlaTyrValIleuAsnLysGlnAlaTyrIle---	278
OY	647	AACATCAGACAGCGAGCGCGGCGCTACCTCTGCCACCAAGCTTATCCAAATGTC	706
Db	279	-----GluLysAsnGlyLysIleIleIleHisArgThrSerSerAsn---	291
OY	707	CTGGATGACCCCAAGTCTGCGGGGTGTGGCCACCTTCGTATACAGAAAGATGTACCCGC	766
Db	292	--SerLysHisIleThrAsnGlyValProSerTyrIleValGlnGlnGluLeuArgGly	310
OY	767	TTTCACTGGGTACTGTGGTGGTCCACACAGCTCTGGAGAGGTTACAGAGGCTCAAGAGC	826
Db	311	PheGlnGlyIleLeuPTrp-----SerLysIser---LysThr	321
OY	827	CTGGCAATCCTGTATGAGGAAGTGCATGATCCGAGGTGAGGTC-----ATT	874
Db	322	---ArgLeuLeuTyrGlnHisValAsnGlnGluLysValAlaGlnSerGlnPheGlyVal	340
OY	875	CAGTCCCTCTCTCTGCGCTAGAAAGAAAGAGACGAGACTGTATGGTATCCCCAGACA	934
Db	341	AsnGlyAspProProValAla-----ProMetLysTyrProArgAla	354
OY	935	GGCAGACAGATCCCAAGATTTGCTTAAACTGGCTGATGCTTCAGACTACAGACCCAGGCG	994
Db	355	GlyThrLysAsnAlaTyrSerThrLeuArgMetValIleLeuLnu-----AsnGly	371
OY	995	AAGATCGCTGACCCAGCGAAGAGACAGTGTGCACAGCCTTCAGCTCGCTTCCGAG	1054
Db	372	LysAlaTyrAspValProLeuLysAspGlnVal-----IleTyrLysHisCysProPhe	389
OY	1055	GTCGAGTACATCCCGAGGCCGGGTGAGACCCGGGATGCGCAATACCGCTGGGCCATGTC	1114
Db	390	TyrGlnTyrIleThrArgAlaGlyPhePheSerAspGlyThrThrValIThrValGlnVal	409
OY	1115	CTGGACCGCCCAACAGATGGTGCCTCAGCTGTCTCTCTCTCCCGCCGCGCTGATCCCG	1174
Db	410	MetSerArgAspGlnAlaGlnCysSerLeuLeuLeuIleProTyrThrAspPheLeuLeu	429
OY	1175	AGCACAGAG-----	1183
Db	430	ProGlnGluLeuGlyGlySerIleLysGluAspAsnLeuGlnLeuSerThrAspLeuAsn	449

QY	1184	-----AATGAGAGACAGCGGCTAAGCCTTGCACAGAGCTGTGCCAAG-----	1223
Db	450	MetGlyValTrrpAspRlySerHisnIsgluThiMetGluLysProProArgGlyLys	469
QY	1226	-----AATGCGACGCGGTATGAGGTGGTGCAGAGAGCTACCAACGCTGGAGTAA	1276
Db	470	LeuArgGlyThrValGln-----IleHisLysAlaArgAsnAspTyrTrpIleAsn	486
QY	1277	GTTATGACATCTTCTATGCCCTTCCCAATAGAGAGGAGAGAGACTGTGC---TTT	1333
Db	487	ThrHisAsnAlaIleTyrTrpIleLysIleThrAspGluGlnHisProMetGlyGluPhe	506
QY	1334	CTCCGCGCCATGATGACAGACCGGCTCTGCATTTGTACAAAGTACAGCGCGTTTA	1393
Db	507	IleTyrCysLeuGluLysProAsnGlySerCys---LeuAlaLeuIleSerAlaGluLeu	525
QY	1394	AAATCCACGGCTACGATTTGGAGTACGCCCTTCAGCCCGGGGAAGATTAATTAAGTC	1453
Db	526	AspGlnAsnGlyTyr-----Cys	531
QY	1454	CCCAATTAAAGAGAGATTGCTGTACACGCGGTGAATGGAGATTGTTGGACGACGCGC	1513
Db	532	ArgHisThrGluGlnLysIleLeuMetAlaGluAsnPheSerIleAsnLysMetGly	551
QY	1514	TCCAAATCTGGGTCAATGAGAGACCAAGCTGTACTTCCAGGCGACCAAGACACAG	1573
Db	552	-----IleValValAspGluValArgGluLeuValTyrTyrValAlaAsnIleHis	569
QY	1574	CCGCTGAGACCAACCTCTACGTGTGTACGTATGAGACGCGCGGAGATGCTACGCTC	1633
Db	570	ProThrGluTrrpAsnIle---CysValSerHisTyrArgThrGluGlnHisAlaGluLeu	588
QY	1634	ACCAAGCGCCGCTCTCCATAGTGCTGCATGAGACAGAC-----	1675
Db	589	ThrLeuSerGlyIle-----CysPheLysSerGluArgAlaAsnGlyLysLeuAla	605
QY	1676	TTCGACATG-----TTCGTGAGCCACTACAGACGCGTGAAGACGCGGCTGCGTG	1726
Db	606	LeuAspLeuAsnHisGlyPheAlaCysTyrMetThrSerValGlySerProAlaGluCys	625
QY	1727	CACGCTACAAAGCTGACGCGCGCGAGACGACGCGCTGCACAAAGACGCGGCTTCTGG	1786
Db	626	ArgPheTyrSerPheArgTrrpLysGluAsnGluValLeu-----ProSerThrValTyr	643
QY	1787	GCTGACATGATGAGGAGGACAGCGAGCTGCC-----CGGAT-----TATGTCCTCA	1834
Db	644	AlaAlaAsnIleThrValSerGlnHisProGluGlnProAsnLeuHisPheSerPro	663
QY	1835	GAGATCTTCATTTCTCCACAG---CGCTCGAGATGTGCGGCTACAGGCAATGATACAG	1891
Db	664	GluMetIleGluPheGlnSerLysLysThrGlyLeuMetHisIleAlaMetIleLeuArg	683
QY	1892	CCCCACGCGCTTGACAGCGAGGAGAGACACCCACGCGCTCTTGTATATGAGGCGCC	1951
Db	684	ProSerAsnPheAspProGlyLysLysTyrProValPheHisTyrValTyrGlyLysPro	703
QY	1952	CAGTGGACGCGGGGATTACTCTCTCAAGAGATCAAGTACTGTGGGCTCAACACTG	2011
Db	704	GlyIleGlnIleValHisAsnAspPheSerTrrpIleGluIleLeuArg-----Phe	720
QY	2012	GCGTCCCGGCGGCTACGCGCGTGTGTATGATGACAGGAGGCGCTCTGACGAGCGCTT	2071
Db	721	CysArgGluGluGlyTrrpValValPheIleAspAsnArgGlySerAlaHisArgGlyIle	740
QY	2072	CGGTTCCAGCGCGCTGAAAACCAATGGCGCAGGTGAGATGAGAGACAGGTGGAG	2131
Db	741	GluPheGluIleValHisIleHisLysLysMetGlyThrValGluValGluLysArgValGlu	760
QY	2132	GCGCTGCGAGCTCGCGCGCGAGAGAT---GCGTTATGAGCTAGCGGATGTCCATC	2188
Db	761	GlyLeuGlnMetLeuAlaGluArgThrGlyGlyPheMetAspMetSerIleArgValVal	780
QY	2189	CATGCGTGGTCTACGGGGGCTCTCTGCTCATGGGCTATATCCAAAGCGCCAGGTG	2248

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 12, 2002, 11:43:57 ; Search time 40.5 Seconds
(without alignments)
12423.890 Million cell updates/sec

Title: US-09-976-674-4
Perfect score: 4795
Sequence: 1 caagcttaccatgcgcaccca.....cttgagcgccgcgagtcgcg 2617

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame-n2p.model -DB=tblh
-Q=/gnu2.1/USPTO.spool/US0997674/runtat_04122002_162400_6009/app_query.fasta_1.2759
-DB=PIR_73 -OFMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.ccd -LIST=45
-DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0997674.GCCN.1.1.44.0.runtat_04122002_162400_6009 -KCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	934.5	19.5	931	2	T32919
2	805.5	16.8	738	2	A87516
3	776	16.2	741	2	JCS142
4	585.5	12.2	711	2	S66261
5	508.5	10.6	793	2	T41703
6	487.5	10.2	766	1	CDH026
7	484.5	10.1	792	1	A39914
8	480	10.0	760	1	S23752
9	479.5	10.0	799	2	T25174
10	467.5	9.7	779	2	T25173
11	452	9.4	865	2	T54331
12	445.5	9.3	803	2	I68600
13	442.5	9.2	803	2	A41793
14	432.5	9.0	818	1	A30107

15	417	8.7	931	2	A49737	dipeptidyl aminope
16	405.5	8.5	795	2	F82858	dipeptidyl-peptida
17	359.5	7.5	1367	1	S48478	glucan 1,4-alpha-g
18	352	7.3	2232	2	T34434	hypothetical prote
19	333.5	7.0	829	2	T19514	hypothetical prote
20	328	6.8	3570	2	T45025	hypothetical prote
21	314	6.5	759	2	I38593	mucin MUC5B, trach
22	302	6.3	1791	2	T02345	fibroblast activat
23	299	6.2	580	2	T02345	hypothetical prote
24	291	6.1	4776	2	T43481	probable mucin DRF
25	286.5	6.0	683	2	E87495	cell wall surface
26	286	6.0	528	2	I47141	prolyl oligopeptid
27	281.5	5.9	3020	2	A43932	gastric mucin (clo
28	280	5.8	1106	2	T00405	mucin 2 precursor,
29	276.5	5.8	924	2	S27923	hypothetical 119.5
30	275.5	5.7	709	2	B82880	gene lrf3 protein -
31	269.5	5.6	2187	2	T30826	alanyl dipeptidyl
32	267	5.5	1414	1	S23809	nascent polypeptid
33	265.5	5.5	660	1	Q0BE3	collagen alpha 2(I
34	263	5.5	660	1	Q0BE3	BHLFI protein - hu
35	262	5.5	743	2	T37700	probable dipeptid
36	256.5	5.3	1032	2	T34433	hypothetical prote
37	255.5	5.3	1460	1	EDBEIF	immediate-early pr
38	254	5.3	1459	2	T32271	hypothetical prote
39	252	5.3	1777	2	T34369	hypothetical prote
40	252	5.3	13288	2	T03099	mucin, submaxillar
41	251.5	5.2	1763	2	S16366	collagen alpha 2(I
42	249.5	5.2	1690	1	CGH01B	collagen alpha 4(I
43	249	5.2	657	2	E70025	probable acylamino
44	248.5	5.2	839	2	F75518	hypothetical prote
45	248	5.2	1151	2	T18535	high molecular mas

ALIGNMENTS

RESULT 1
T32919
hypothetical protein K02F2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32919
R:Magdi, L.; Goela, D.
A:Description: The sequence of C. elegans cosmid K02F2.
A:Reference number: 221246
A:Accession: T32919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-931 <MAG>
A:Cross-references: EMBL:AF043699; PIDN:AA897564.1; GSPDB:GN00019; CESP:K02F2.1
A:Experimental source: strain Bristol N2; clone K02F2
C:Genetics:
A:Gene: CESP:K02F2.1
A:Map position: 1
A:Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2;

Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
		3.47e-48	934.50	45.368	29.508	2
				Mismatches:		
				Indels:		
				Gaps:		33
US-09-976-674-4 (1-2617) x T32919 (1-931)						
QY 224 CTCTACCTACCTGGGAATGCCATATGCGCAGAGACTCCCTCTACTCTGAGATT 283						
DB 82 MetTyrAlaIleSerSerValProGlyThrAsnThcInSerIlePheSerValThrIle 101						
QY 284 CCCAAGAGGTC-----CGGAAGAGGCTCTGCTGCTCTG 319						
DB 102 ProLeuGluIleuValGIuIySalGlnValAlaIaAspArgIySrhGluIleuLysIleuLys 121						

Run on: December, 12, 2002, 10:36:52 ; Search time 55 Seconds

12680.614 Million cell updates/sec

Title: US-09-976-674-4

Sequence: 1 caagcttaccatgcca...tctgagcgccgcgatccg 2617

Scoring table: BLOSUM62

Xgapop	10.0	, Xgapext	0.5
Ygapop	10.0	, Ygapext	0.5
Fgapop	6.0	, Fgapext	7.0
Delop	6.0	, Delext	7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

MODEL=irama_n2p_model -DEV=xlh
-O/-cgrn1_US/SPOT.spool/US099976674/runat.04122002.162358.5964/app_query.fasta_1.2755
-DBA=Genesec.101002 -QPM=fastlan -SUFFIX=rag -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human0.cdi
-LIST=45 -DOCALLIG=200 -THR SCORE=pct -THR MAX=10 -THR MIN=0 -ALLIG=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=5E+50 -MINLEN=0 -MAXLEN=20000000000
-USER=US09976674.cgc@1.1.61 -runat.04122002.162358.5964 -NCP=6 -ICP=3
-NO_XLXPY -NO_MAMP -LAREQUEST=NEG_SCORES=0 -WAT -LONGLOG -DEV_TIMEOUT=120
-WARL_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP -DEV_TIMEOUT=7
YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : A_Geneseq_101002:*

1:	/SID2/gcgdata/genseq/genseqp-emb1/A11980.DAT *
2:	/SID2/gcgdata/genseq/genseqp-emb1/A11981.DAT *
3:	/SID2/gcgdata/genseq/genseqp-emb1/A11982.DAT *
4:	/SID2/gcgdata/genseq/genseqp-emb1/A11983.DAT *
5:	/SID2/gcgdata/genseq/genseqp-emb1/A11984.DAT *
6:	/SID2/gcgdata/genseq/genseqp-emb1/A11985.DAT *
7:	/SID2/gcgdata/genseq/genseqp-emb1/A11986.DAT *
8:	/SID2/gcgdata/genseq/genseqp-emb1/A11987.DAT *
9:	/SID2/gcgdata/genseq/genseqp-emb1/A11988.DAT *
10:	/SID2/gcgdata/genseq/genseqp-emb1/A11989.DAT *
11:	/SID2/gcgdata/genseq/genseqp-emb1/A11990.DAT *
12:	/SID2/gcgdata/genseq/genseqp-emb1/A11991.DAT *
13:	/SID2/gcgdata/genseq/genseqp-emb1/A11992.DAT *
14:	/SID2/gcgdata/genseq/genseqp-emb1/A11993.DAT *
15:	/SID2/gcgdata/genseq/genseqp-emb1/A11994.DAT *
16:	/SID2/gcgdata/genseq/genseqp-emb1/A11995.DAT *
17:	/SID2/gcgdata/genseq/genseqp-emb1/A11996.DAT *
18:	/SID2/gcgdata/genseq/genseqp-emb1/A11997.DAT *
19:	/SID2/gcgdata/genseq/genseqp-emb1/A11998.DAT *
20:	/SID2/gcgdata/genseq/genseqp-emb1/A11999.DAT *
21:	/SID2/gcgdata/genseq/genseqp-emb1/A12000.DAT *
22:	/SID2/gcgdata/genseq/genseqp-emb1/A12001.DAT *
23:	/SID2/gcgdata/genseq/genseqp-emb1/A12002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Query Score	Length	DB	ID	Description
---------------	----------------	--------	----	----	-------------

1	4646	96.9	863	23	ABG61592	Human DPPIV related
2	4646	96.9	892	23	ABG61602	Human DPPR-2 splicd
3	4646	96.9	892	23	ABG61604	Human DPPR-2 splicd
4	4636	96.7	969	23	AAE24168	Human dipeptidyl p
5	4558.5	95.1	879	23	ABG61607	Human DPPR-2 splicd
6	4558.5	95.1	879	23	ABG61608	Human DPPR-2 splicd
7	4458	93.0	830	23	AAE24171	Human dipeptidyl p
8	4279	89.2	869	23	AAE24169	Alternative versio
9	4259.5	88.8	832	23	ABG61605	Human DPPR-2 splicd
10	4259.5	88.8	832	23	ABG61606	Human DPPR-2 splicd
11	4172	87.0	819	23	ABG61609	Human DPPR-2 splicd
12	4172	87.0	819	23	ABG61610	Human DPPR-2 splicd
13	4129	86.1	847	23	AAE23875	Murine dipeptidyl
14	3817.5	79.6	737	22	AAE38724	Human polypeptide
15	3539.5	73.8	623	22	AAAM4510	Human polypeptide
16	3409.5	71.1	720	21	AAAB41626	Human ORFX ORF1390
17	2870	59.9	882	22	AAAB47187	Human DPP8. Homo
18	2870	59.9	882	23	AAE24170	Human dipeptidyl p
19	2870	59.9	882	23	ABG61591	Human DPPIV related
20	2870	59.9	882	23	AAU74749	Human protease PR
21	2870	59.9	882	23	AAAG78415	Amino acid sequenc
22	2688	56.1	497	23	ABG64615	Human albumin fusi
23	2688	56.1	497	23	AAU96169	Human secreted prote
24	2684	56.0	497	23	ABG64844	Human albumin fusi
25	2684	56.0	497	23	AAU96192	Human secreted prote
26	2572	53.6	508	21	AAAB42928	Human ORFX ORF2699
27	2562	53.4	518	21	ABG61603	Human DPPR-2 splicd
28	2547	53.1	518	21	AAAY90299	Human peptidase, h
29	2414	50.3	782	23	ABBB97361	Novel human protease
30	2236.5	46.6	512	22	AAAM40526	Human polypeptide
31	2215.5	46.2	460	22	AAAM37362	Human polypeptide
32	2178	45.4	724	23	ABBB97362	Novel human protease
33	2137.5	44.6	632	22	AAAB95565	Human protein sequ
34	2001	41.7	690	23	ABG61594	Human DPPR-1 splicd
35	1999.5	41.7	661	23	ABG61596	Human DPPR-1 splicd
36	1997	41.6	658	23	ABG61600	Human DPPR-1 splicd
37	1816	37.9	613	23	ABG61601	Human DPPR-1 splicd
38	1631.5	34.0	1042	22	ABBB60137	Drosophila melanog
39	1631.5	34.0	1102	22	ABBB62029	Drosophila melanog
40	1599.5	33.4	580	22	AAAE14337	Human protease PR
41	1494.5	31.2	465	22	AAAB47189	Human DPP8 318Thr
42	1250.5	25.5	360	22	AAAB47190	Human DPP8 244Glu
43	1093.5	22.8	310	22	AAAB47188	Human DPP8 524Phe
44	1093.5	22.8	310	23	ABBB08994	Human dipeptidyl p
45	1067.5	22.3	349	23	ABG64842	Human albumin fusi

ALIGNMENTS

RESULT 1

ID ABG61592 standard; Protein; 863 AA.

AC ABG61592;

DT 12-AUG-2002 (first entry)

DE Human DPPIV related serine protease DPPR-2

KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR

KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke

KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;

KW metabolic disorder.

05. Homo sapiens.

XX WO200231134-A2.
 XX 18-APR-2002.
 PD
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI; 2002-444178/47.
 DR N-PSDB; ABK83323.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 XX
 PS
 XX
 XX Claim 17; Fig 1; 113pp: English.

CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPIV)-related
 CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
 CC
 XX
 XX

SQ Sequence 863 AA;

Alignment Scores:

Pred. No.: 0 Length: 863
 Score: 4646.00 Matches: 863
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.89% Indels: 0
 DB: 23 Gaps: 0

US-09-976-674-4 (1-2617) x ABG61592 (1-863)

QY 11 ATGGCCACACCGGACCCCAACGCGGACGAGCGAGCGCCGACAGATGACCG 70
 Db 1 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAsp 20
 QY 71 GCGCGCCGCTTCCAGGTGACAGACACTGTTGGAGCGGCTCCGAGCATCATCAGGC 130
 Db 21 AlaAlaArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 40
 QY 131 AGCGGCAAGTACGCGGCTCATGTCAACAAGGCGCCCAACGACTCCAGTTGGTCAG 190
 Db 41 SerArgLysLysSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 60
 QY 191 AAGACGATGAGTCTGGGCGCCACATCCACGCGCTTACTACTAGGAGATGCGATATGCG 250
 Db 61 LysThrAspGlnSerGlyProHisSerHisArgLeuLysTrpLeuGlyMetProTyrGly 80
 QY 251 AGCGAAGAACTCCCTCTCTACTGTGAGATTCCCAAGAGGTCCGGAAGAGGCTCTG 310
 Db 81 SerArgLysLysSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 100
 QY 311 CTGCTCTGCTCGGAGGAGATGCTGATCATTTCCAGGCGACGCGCCACCATGGGCTC 370
 Db 101 LeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 120

QY 371 TACTCTGGAGAGAGACTGCTGAGGAGCGAAGACGCTGGGGGTCTTCGACATCAC 430
 Db 121 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThr 140
 QY 431 TCCTACACTTTCACACGAGAGTGGCTCTCTCTTCCAGGCGACCAACAGCTTCTC 490
 Db 141 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 160
 QY 491 CACTGTGGCGAGGCGGACAGAAAGGCTCATGGTGCCTATGAACCGCTGGAATC 550
 Db 161 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 180
 QY 551 AAGACCGATGCTCAGGCGCCCGGATGAGACCCCAAAATCTGCCCTGCGACCTGCTTC 610
 Db 181 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 200
 QY 611 TTCTCTTCATCAATAACAGGACCTGTGGTGCCCAACATCGAGAGCGGAGCGG 670
 Db 201 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnThrGlyGlnGluArg 220

QY 671 CGGCTGACCTTTCGCGCAAGGTTTATCCAAATCCTGATGACCCCAAGTCTGGGCT 730

Db 221 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 240

QY 731 GTGCGCACCTTCGTCATACAGAGAGAGTTCGACCGCTTCACTGGTACTGGTGGCCC 790

Db 241 ValAlaThrPheValIleGlnGlnGlnLysPheAspArgPheThrGlyTrpTrpCysPro 260

QY 791 ACACGCTCTGGGAGGTTTCAGAGGCGCTCAACAGCTGCGAATCTGTATGAGAAATC 850

Db 261 ThrIleSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGluVal 280

QY 851 GATGAGTCCGAGGAGGATCATTCAGTCCCTCTGCGTGAAGAAAGGAGAGC 910

Db 281 AspLeuSerGlnValGlnValIleHisValProSerProAlaLeuGlnGluValGlyThr 300

QY 911 GACTCTATCGGTACCCACAGACAGCGACAGAAATCCAGATGCTTGAATGAGTCT 970

Db 301 AspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 320

QY 971 GAGTTCAGACTGACACCGAGGAGACATGCTCTCCAGCCAGAGAGAGAGCTGTCCAG 1030

Db 321 GlnPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGln 340

QY 1031 CCCCTGAGCTGCGGTCCGAGGAGGTGAGATGCGCCAGGCGGCTGAGCCCGGAT 1090

Db 341 ProPheSerSerLeuPheProLysValGlnIleValIleAlaArgAlaGlyTrpThrArgAsp 360

QY 1091 GGCAAATACGCTGGGACATGTTCTGGACCGGCGCCAGGAGTGCATCAGCTGTCCTC 1150

Db 361 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 380

QY 1151 CTCGCGCGGCGCTGTTCAATCCGAGACAGAGAAATGAGAGAGCGGCTGTGCC 1210

Db 381 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnGlnAlaGlnAlaSerAla 400

QY 1211 AGAGCTGTCCCGAGGATGTCAGCGCTATGTGTGACGAGAGGTCCACCAACGTTGG 1270

Db 401 ArgAlaValAlaProAlaGlnValGlnProTyrValValTyrGlnGlnValIleHisValTrp 420

QY 1271 ATCAATGTCATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGAGAGCTCTGC 1330

Db 421 IleAsnValIleHisAspIlePheTyrProPheProGlnSerGlnGlyGlnAspGlnLeuCys 440

QY 1331 TTCTCTCGCGCCATGATGATCAGAGCGGCTTCTGCCATTTGTACAAAGTACCGCGCTT 1390

Db 441 PheLeuArgAlaAsnGlnLysLysThrGlyPheCysHisLeuTyrLysValIleAlaVal 460

QY 1391 TTTAAATCCAGGCGCTGAGTGTGAGTGAAGCCCTTCAGCGCGGAGAGATGAAATTAAG 1450

Db 461 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlyAspGlnPheLys 480

```

QY 1451 TGCCCATTAAGAGAGATTGCTCTGACACAGCGGTGAATGAGAGTTTGGACGAGC 1510
    |||||||
Db 481 CysProIleuysgluGluIleAlaIleuThrSerIleYlurPgluValIleuAlaArgHis 500
QY 1511 GGCCTCAGAGATCTGGGTCAATGAGAGACCAAGCTGGTACTTCACAGGACCAAGAGAC 1570
    |||||||
Db 501 GlyserysIleerPvalasngluGluThrIlysleuValIYrPheglnglYThrIlysasp 520
QY 1571 ACGCGCGGAGACACACCTTACGTGGTGCATGATGAGGCGGCGGAGACATGTCGC 1630
    |||||||
Db 521 ThrProIleuGluHisIleuIYrValValSerIYrGluAlaIlaGluIleValArg 540
QY 1631 CTCACACAGCGCGCTTCTCCCATAGCTGCTCAAGACCAAGCTTCGACATGTCGTC 1690
    |||||||
Db 541 LeuThrThrProglIYrPheSerHisSerCysSerMetSerGluAnanPheaspMetPheVal 560
QY 1691 AGCCACTACAGACAGCGTGAAGACGCGCGCTGCGTGCACGTCTACAGACTGACGCGCC 1750
    |||||||
Db 561 SerHisIYrSerSerValSerThrProIYrCysValHisValIYrIYlsleuSerGlyPro 580
QY 1751 GAGACGACCGCCGACACACAGCGCGCTTGGGCTGATACATGATGAGAGACGACGAC 1810
    |||||||
Db 581 AspAspAspProIleuHisIleYsGluProArgPheIYrPalasSerMetGluAlaIaIaSer 600
QY 1811 TGCCCGCGGATATGTTCTCCAGAGATCTTCATTTCCACAGCGCGCTCGATGTCGG 1870
    |||||||
Db 601 CysProIYrAspIYrValIProProgluIlePheHisSerHisThrArgSerAspValArg 620
QY 1871 CTCACAGGATGATCTACACAGCGCGCTGACAGCGGAGGAGAAAGACCGCACCGCTC 1930
    |||||||
Db 621 LeuIYrGluMetIleIYrIYrIYsProHisAlaIeugInProglIYsIYsHisProThrVal 640
QY 1931 CTCCTTGATATGAGAGCGCGCCAGGTGAGTGTGAATCTGCTTAAAGCATCAAG 1990
    |||||||
Db 641 LeuPheValIYrGluGlyIYrProglInValGluIleuValaIasnaSerPheIYsGlyIleYs 660
QY 1991 TACTTGGCGCTCAACACACAGCGCGCTGAGCTACGCGGCTGATGATGAGCGGCGAG 2050
    |||||||
Db 661 TyIleuArgIleuAsnThrIleuAlaSerIleuGlyIYrIaValaIValaIleuSpIYArg 680
QY 2051 GGCCTCTGTCAGCAGGCGCTTCGTCAGAGGCGCGCTGAAACCAAAATGGGCGCAGT 2110
    |||||||
Db 681 GlySerCysGluArgGlyLeuArgPheGluGlyAlaIleuIYsAsnGluMetGlyGluVal 700
QY 2111 GAGATCGAGGACACAGGTGAGGCGCTGCAATTCGTGCGCGGAGCATATGCTTCATCAC 2170
    |||||||
Db 701 GluIleGluAspGluValGluGlyLeuGluIleuValaIaGluIYsIYrGlyPheIleasp 720
QY 2171 CTGAGCGGAGTGGCATGCTGCGTGGTCTACGCGGCGCTTCCTGCTCATGGCGGCTA 2230
    |||||||
Db 721 LeuSerArgValaIaIleHisGlyIYrPserIYrGlyIYrPheIleuSerIleuMetGlyLeu 740
QY 2231 ATCCACAAAGCGCGGTGTCAGGTGCGCATCGCGGCGTCCCGGTACCGCTGTGATG 2290
    |||||||
Db 741 IleHisIYsProglInValPheIYsValaIaIleAlaGlyAlaIProValThrValTrpMet 760
QY 2291 GCGTACGACACAGGCTACACTGACGCGCTACATGACGCTCCCTGAGACACACGACGCGC 2350
    |||||||
Db 761 AlaIYrAspIYrThrGlyIYrThrGluArgIYrMetAspValIProgluAsnAsnGluHisGly 780
QY 2351 TATGAGGCGGCTTCGTCGCGCTGACGTCGAGAGAGCTGCCCATGAGCCACACCGCTG 2410
    |||||||
Db 781 TyIYrGluAlaGlySerValaIaIleuHisValaIgluIYsIleuProAsnGluIYrPheLeu 800
QY 2411 CTTATCTCTCACGCGCTTCCTGAGCAAAAGCTGCACTTTTCCACACAACCTTCCTGTC 2470
    |||||||
Db 801 LeuIleIleuHisGlyPheIleuAspGluAsnValaIHisPhePheHisThrAsnPheIleVal 820
QY 2471 TCCCAACTGATCCGACGACGAGGAAACCTTACCAAGCTCCGATCTACCCCAAGAGAGAC 2530
    |||||||
Db 821 SerGluIleuIleuArgIaGlyIYsProIYrGluIleuGluIleIYrProAsnGluIaArgHis 840
QY 2531 AGTATTCGCTGCGCGGAGTCCGGGAGAGACATATGAAGTACAGTTCGCTGCTCTACAG 2590

```

```

Db 841 SerIleArgCysProGluSerGlyGluHisIYrGluValThrIleuHisPheIleuGln 860
QY 2591 GAATACCTC 2599
    |||||||
Db 861 GluTyIleu 863

RESULT 2
ABG61602
ID ABG61602 standard; Protein: 892 AA.
AC ABG61602:
DT 12-AUG-2002 (first entry)
DE Human DPPR-2 splice variant #1.
XX
XX Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
PN W020031134-A2.
XX
XX 18-APR-2002.
PD
XX
XX 12-OCT-2001; 2001MO-US11874.
PE
XX 12-OCT-2001; 2000US-240117P.
XX
XX 12-OCT-2000; 2000US-240117P.
PR
XX
XX (FERR ) FERRING BV.
PA
XX
XX Qi S, Aktinsanya KO, Riviere PJ, Junien J;
PI WPI: 2002-444178/47.
XX
XX N-PSDB; ABK83333.
DR
XX
XX New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
XX
XX Disclosure; Page 76-78; 113pp; English.
PS
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPPR proteins.
XX
XX
SQ Sequence 892 AA;

```

```

Alignment Scores:
Pred. No.: 0
Score: 4646.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 96.89%
DB: 23
Gaps: 0

```

US-09-976-674-4 (1-2617) x ABG61602 (1-892)

QY 11 ATGGCCACGACCGGAGCCGACGAGCGAGCGAGCGGCGGACGATGACCG 70
Db 30 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaThrAspAspPro 49
QY 71 GCGCGCCGCTTCCAGGTGAGAAACACTCGTGGAGCGGCTCCGAGCATCATCACGGC 130
Db 50 AlaAlaArgPheGlnValGlnLysHisSerTrpAspArgLeuArgSerIleIleHisGly 69
QY 131 AGCGGCAAGTACTGGGCTTCATTGTTCACAAGCGGCCCGCCACGATTCACGTTTGTCAG 190
Db 70 SerArgLysTrpSerGlyLeuIleValAlaSnLysAlaProHisAspPheGlnPheValGln 89
QY 191 AGACGGATAGTGTGGGCGCCACTCCACCGGCTCTACTACCTGGGGAAGGCCATATGGC 250
Db 90 LysThrAspLysSerGlyProHisSerHisArgLeuTrpTyrLeuGlyMetProTyrGly 109
QY 251 AGCGGAGAGAACTCCCTCTCTACTGTAGATTCACAAAGAGTCCGAAAGAGCTCTG 310
Db 110 SerArgGlnAsnSerLeuLeuTrpSerGluIleProLysLysValArgLysGluAlaLeu 129
QY 311 CTGCTCTGTCTGGAGAGAGATGCTGATCATTTCCAGGCGCACGCCCGCCACATGGGCTC 370
Db 130 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
QY 371 TACTCTGGAGAGAGAGAGTGTGAGGAGCGGAAACGCTGGGGGTCTTCGCGCATCAC 430
Db 150 TyrSerArgGlnGluGlnLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThr 169
QY 431 TCCTAGACATTCACAGCGAGAGTGGCGCTCTCTCTTCCAGGCGCAACAGCGCTTTC 490
Db 170 SerLysAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
QY 491 CACTGTGGCAAGCGGCGCAGAAAGCGCTCATGCTGTCCTCCATGAAGACCGCTGGAATC 550
Db 190 HisLysTrpAspArgLysGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
QY 551 AAGACGAGTGTCTGAGGCGCCGCGATGAGCCCAAAATGTGCCCTGGCGACCTGCTTC 610
Db 210 LysThrGlnLysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
QY 611 TTCTCTTCAATTAACAGCGAGCTGTGGTGCGCAACATCGAGACAGCGGAGAGCGG 670
Db 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlnGluArg 249
QY 671 CGGCTGACCTTGTGGCCAAAGGTTTATCCATGCTCGATGCCAGTCCGCGGGT 730
Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
QY 731 GTGGCCACTTTCATACAGAAAGATTCGACGCTTCACATGGGTACTGGTGGTGGCCC 790
Db 270 ValAlaThrPheValIleGlnGlnGluPheAspArgPheThrGlyTyrTrpTrpCysPro 289
QY 791 ACAAGCTCTGGGAAGGTTCAAGAGCGCTCAAGAGCTGCGAATCTGTATGAGGAATC 850
Db 290 ThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGluVal 309
QY 851 GATGAGTCCAGGTGAGGTGATTCAGTCCCTCTCTGCTCGCTAGAAAGAAAGAAAGACG 910
Db 310 AspLysSerGlnValGlnValIleHisValProSerProAlaLeuGlnGluArgLysThr 329
QY 911 GACTCGATGCGTACCCAGAGACAGGACAGAAATCCCAAGATTTGGCTTGAATGCTGCT 970
Db 330 AspSerTrpArgLysTrpArgGlyThrGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
QY 971 GAGTTCAGACTGACAGCGAGGAGAGATGCTTCAAGCCAGGAAAGAGAGCTGGTCCAG 1030
Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnLysGlnLeuValGln 369
QY 1031 CCTTTCAGCTCGCTGTTCCGGAAGGTGAGTACATGCCCAAGGCGCGGCTGACCCGGCAT 1090
|||||

Db 370 ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGluTyrTrpThrArgAsp 389
QY 1091 GGCAAATATACCTGTGGGCTATGTTCTGTGACGGGCCAGGACAGTGGCTCAAGTGTCTC 1150
Db 390 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 409
QY 1151 CTCGCCCGGCGCTGTTCATCCGAGACAGAGAAATGAGAGACAGCGGCTAGCCTGGCC 1210
Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnArgLeuAlaSerAla 429
QY 1211 AGAGCTCCCCAGAGATGTCCAGCCGTATGTGTGACAGAGAGCTCACCACAGTCTGG 1270
Db 430 ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGluValIleThrAsnValTrp 449
QY 1271 ATCAATGTCATGACATCTTCTATCCCTTCCCAATCAAGAGGAGAGAGAGCTCTGC 1330
Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlnAspGlnLeuCys 469
QY 1331 TTTCTCGCGCCCAATGAAATGAAACAGCGGCTTGCACATTTGTACAAAGTCCACGCCGT 1390
Db 470 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValIleThrAlaVal 489
QY 1391 TTAATAATCCAGGCTACGATGAGTGAAGAGCGCTTCCAGCCCGGGAAGATTAATTAG 1450
Db 490 LeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProGlnGlnAspGluPheLys 509
QY 1451 TGCCCCATTAAGAGAGATGTGCTGACAGCGGTAATGGAGGAGTTTGGCGAGGCAC 1510
Db 510 CysProIleLysGlnGluIleAlaLeuThrSerGlyIleTrpGlnValLeuAlaArgHis 529
QY 1511 GGCTCCAAATCTGGGTCAATGAGAGACAAAGCTGTGTACTTCCAGGCGACCAAGAC 1570
Db 530 GlySerLysIleTrpValaAsnGlnGluThrLysLeuValTyrPheGlnGlnTyrLysAsp 549
QY 1571 AGCGCGCTGAGACACACCTCTAGCTGAGTACGAGGAGGCGCGGAGATCGTACGC 1630
Db 550 ThrProLeuGlnHisHisLeuTyrValValSerTyrGlnAlaAlaGlyGluIleValArg 569
QY 1631 CTCACAGCGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTCGACATGTTGCTC 1690
Db 570 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 589
QY 1691 AGCCACTACAGACGTGAGACAGCGCGCGCTGCGTGCACGTTCACAGCTGAGCGGCC 1750
Db 590 SerHisTyrSerSerValSerThrProCysValHisValTyrLysLeuSerGlyPro 609
QY 1751 GAGCAGACGCCCTGCAAGAGAGAGCGCGCTGCGGTGAGTGAAGAGAGAGCGCCAGC 1810
Db 610 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaAlaSer 629
QY 1811 TGCCCCCGGATTAATGTTCTCCAGAGATCTTCATTTCCACAGCGCGCTGGATGTGCG 1870
Db 630 CysProProAspArgValProProGluIlePheHisPheHisThrArgSerArgValArg 649
QY 1871 CTCTAGCGCATGATCTCAAGCCCGACGCGCTTGACGCGAGGAAAGACACCCACCGTC 1930
Db 650 LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 669
QY 1931 CTCTTGTATATGAGAGCGCCAGGCTGACGCTGGTAATAATCTTCAAGGCATCAAG 1990
Db 670 LeuPheValTyrGlnGlyProGlnValGlnLeuValaAsnAsnSerPheLysGlyIleLys 689
QY 1991 TACTTGGCGCTCAACACACTGCGGCTCCCTGGGTACGCGCTGTGTGATGAGCGGACG 2050
Db 690 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValaIleAspGlyArg 709
QY 2051 GCGTCTGTACGAGGCGCTTGGTTGGAAGGGCGCTGAATAACAAATGGCGCAGGTG 2110
Db 710 GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnVal 729
QY 2111 GAAATCAGAGACAGAGTGGAGGCGCTGCAATGCTGGCGCAGAAATTTGGCTTCATCGAC 2170
Db 730 GluIleGlnAspGlnValGlnGlnLeuGlnPheValaIleGlnLysTyrGlnPheIleAsp 749
|||||

QY 2171 CTGAGCCGAGTTGCGATCATGCGTGGGCTTCAGCGGGGCTTCCTGCGTCATGGGGCTA 2230
 Db 750 LeuSerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 769
 QY 2231 ATCCACAGACCCCGAGTGTTCAGGTGGCCATCGCGGGTGGCCCGGTCCACCGTGTGATG 2290
 Db 770 IleHisLysProGlnValPheLysValAlaIleLeuGlyAlaProValThrValTrpMet 789
 QY 2291 GCGTACGACAGCGGTACACTGACGCGGTACATGACGTCCTTGAAGAACACACGACGCGC 2350
 Db 790 AlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGlnAsnAsnGlnHisGly 809
 QY 2351 TATAGAGCGGGTCCGCGGCGCCCTGACGTGAGAAAGCCCAATGAGCCCAACCGCTTG 2410
 Db 810 TyrGlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeu 829
 QY 2411 CTATACCTCCAGGCTTCCTGAGCAAAACGTGACATTTTCCACAACTTCCTGCTC 2470
 Db 830 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 849
 QY 2471 TCCCACTGATCGACGAGGAAACCTTACAGCTCCAGATCTACCCCAAGACAGACAC 2530
 Db 850 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 869
 QY 2531 AGTATTCGCTCCCGGAGTCCGGGAGACACTATGACGTGCTGCTGCACTTTCTACAG 2590
 Db 870 SerIleArgCysProGlnSerGlyLuhHisTyrGlnValThrLeuLeuHisPheLeuGln 889
 QY 2591 GAATACCTC 2599
 Db 890 GlutTyrLeu 892
 RESULT 3
 ABG61604
 ID ABG61604 standard; Protein; 892 AA.
 AC ABG61604;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPPP-2 splice variant #3.
 XX
 KW Human: serine protease; dipeptidyl peptidase IV-related protein; DPPP;
 KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 OS
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 DR N-PSDB; ABK83335.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT
 XX

PS Disclosure; Page 81-84; 113pp; English.
 XX
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.
 XX
 SQ Sequence 892 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 892
 Score: 4646.00 Matches: 863
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.89% Indels: 0
 DB: 23 Gaps: 0
 US-09-976-674-4 (1-2617) x ABG61604 (1-892)
 QY 11 ATGGCAGCACCGGAGACCCCAAGCGCGACGAGCGAGCGCCGACAGATGACCGG 70
 Db 30 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaIleAlaThrAspAspPro 49
 QY 71 GCGCGCGCTTCAGGTGACAGACACTGTGGAGCGGCTCCGAGCATATCCAGCGC 130
 Db 50 AlaIleArgPheGlnValGlnLysHisSerTrpAspGlyLeuArgSerIleIleHisGly 69
 QY 131 AGCCGCAAGTACTCGGCGCTCATGTGTCAACAAGCGCGCCCAAGCTTCCATTTGTGAG 190
 Db 70 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 89
 QY 191 AAGACGATGAGTGTGGCGCCCACTCCACCGCCTCTACTGAGTCCAGGACCGCCACCATGAGC 250
 Db 90 LysThrAspGlnSerGlyProHisSerHisArgLeuTyrTyrLeuGlnGlyMetProTyrGly 109
 QY 251 AGCCGAGAACTCCCTCTCTACTGTGAGATTCGCAAGAGAGTCCGAAAGAGGCTCTG 310
 Db 110 SerArgGlnAsnSerLeuLeuTyrSerGlnIleProLysValArgLysGlnAlaLeu 129
 QY 311 CTGCTCCCTGCTCGAAGCAATGCTGATCATTTCCAGGCGACCGCCACCATGAGGCTC 370
 Db 130 LeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaIleProHisHisGlyVal 149
 QY 371 TACTCTGGAGAGGAGACTGCTGTGAGGAGGAAAGCTGGGGGCTTGGGATTCACC 430
 Db 150 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlnGlyValPheGlyIleThr 169
 QY 431 TCTACGACTTCCACAGGAGAGTGGCTTCTCTTCCAGGCGACGACAGGCTCTTC 490
 Db 170 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
 QY 491 CACTGTCCGAGCGGCGGCAAGAGAGGCTTCATGGTGTCCCTATGAAACCGCTGGAATC 550
 Db 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
 QY 551 AAGACCCAGTGTGAGGCGCCGATGGAGCCCAAAATCTGCGCTGCGGACCTGCTTC 610
 Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
 QY 611 TTCTCTTCATCATATACAGGACCTGTGGGTGGCCCAATGACAGACGCGAGAGCGG 670
 Db 230 PheSerPheIleAsnAsnSerAspLeuTyrValAlaAsnIleGlnThrGlyGlnGlnArg 249

QY	671	CGGCTGACCTTCTGCACCAAGTTTATCCAAATGTCTGTGATGACCCCAAGTCTGGGGT	730
Db	250	ArgLeuThrPheCysHisGInGlyLeuSerAsnValLeuAspProIysSerIAsGly	269
QY	731	GGGGCCACTTCGTCATACAGGAAGATTGGACGGTCTACGTGGGACAGCGTGGGGCC	790
Db	270	ValAlaIrrPheValIleGInGInGInPheAspArgPheThrGlyTrpTrpIysPro	289
QY	791	ACAGCTCTCCGGGAAGGTTCCAGAGGGCTTCAGACGCTGGGAATCCTGTATAGAGAGTC	850
Db	290	ThrAlaSerTrpGInGlySerGInGlyLeuLysThrLeuArgIleLeuTrpGInGlyVal	309
QY	851	GATGAGTCCGAGGTGGAGGTCATTACAGTCCCTCTCTCTGCTAGAGAAAGAAAGAC	910
Db	310	AspGInSerGInValGInValIleHisValProSerProAlaLeuGInGInIrrArgStr	329
QY	911	GACCTGTTCGGTACCACGAGACGGGAGGAAGAATCCCAAGTTGCCCTTGAAGCTGGT	970
Db	330	AspSerTrpArgTrpProArgThrGlySerLysAsnProIysIleAlaLeuLysLeuAla	349
QY	971	GAGTTCGAGACTGCACAGCCAGGGCAGAGATGCTCTGCACCAGAGAGAAAGAGCTGGTGC	1030
Db	350	GInPheGInThrAspSerGInGlyLysIleAlaSerThrGInGInLysGInLeuValGIn	369
QY	1031	CCCTTCAGCTGCTGTTCOCGAGAGGTGGATACATGCCACGAGCGGGGTGACCCGGAT	1090
Db	370	ProPheSerSerLeuPheProLysValGInTyrIleAlaArgAlaGlyTrpPheArgAsp	389
QY	1091	GGCAATACGGCTGGGCCATGTTCCCTGGACGGGCCACGAGAGTGGTCCAGCTGCTCTC	1150
Db	390	GlyLysTrpAlaTrpAlaMetPheLeuAspArgProGInIrrPheLeuIleuValIleu	409
QY	1151	CTCCCCCGGCGCTTCATCCCGACGACAGAAATAGAGAGACGCGGTACCTCTGCG	1210
Db	410	LeuProProAlaLeuPheIleProSerThrGInAsnGInGInIrrArgLeuAlaSerAla	429
QY	1211	AGAGCTGCCCCAGGAATGTCCACCGCTATGTGTGTACGAGAGGTCACCAAGCTGTGG	1270
Db	430	ArgAlaValProArgAsnValGInProTyrValIValItyGInGInValIrrAsnValIrr	449
QY	1271	ATCATGTTCATGACATCTTCTATCCCTTCCCCCATCAGAGGGAGAGACAGACTCTGC	1330
Db	450	IleAsnValHisAspIlePheTyrProPheProGInSerGInGInGInLysAspIleuLys	469
QY	1331	TTTTCCCGCGCATGATGAATCAAGACACCGGCTTGGCATTTGTACAAATCAACCCGCT	1390
Db	470	PheLeuArgAlaAsnGInLysItyThrGlyPheCysHisIleuTyrItyValIrrAlaVal	489
QY	1391	TTAAAAATCCAGGGCTACGATGTGAGTAGCCCTTGACGCCCGGGAGAGATGAATTAA	1450
Db	490	LeuLysSerGInGlyTrpAspTrpSerGInProPheSerProGlyLysLysPheLys	509
QY	1451	TGCCCATTTAAGGAAGAGATTGCTCTGACACGGGTGAATGGGAGGTTTGGCGAGGCAC	1510
Db	510	CysProIleLysGInGInIleAlaLeuThrSerGlyLutProIleValLeuAlaIrrHis	529
QY	1511	GGCTCAAGATCTGGTCAATGAGAGAGACCAACCTGGTACTTCACGAGGCACCAAGAC	1570
Db	530	GlySerItyIleTrpValAsnGInGInIrrLysLeuValTyrPheGInGInIrrLysAsp	549
QY	1571	ACGGCGGTGAGACACCACTTACGTGTGACCTATGAGCGGGCGCGAGATCTGTAC	1630
Db	550	ThrProLeuGInHisHisLeuTyrValIValSerTyrGInAlaAlaGlyLysIleValIrr	569
QY	1631	CTCACAACGCCGGCTTCTCCATAGCTGCTCATAGCCAGAATCTTGACATGTTGCTC	1690
Db	570	LeuThrTrpProGlyPheSerHisSerCysSerMetSerGInAsnPheAspMetPheVal	589
QY	1691	AGCCACTACACACCGAGAGACGGCGCCCTGGCTGCACAGTCAACAGGTGAGGGGGCC	1750
Db	590	SerHisTyrSerSerValSerThrProProCysValHisValItyLysLeuSerGlyPro	609
QY	1751	GACGACGACCCCTGCACAAAGACGCCCGCTTGGGGCTAGCATATGGAGGACACCGAC	1810

Db	610	ASPARSRPROLEUHNISLVSGLINPROARGRHEPTRALASERMETGLIALALASER	629
QY	1811	TECCCCCGGATTTATGTTTCCSTCCAGACATGTCATTTCCACACGGGCTGGATGGCGG	1870
Db	630	CYSPTROPASRYTVVALPROPTGGLIILPHEHNISPHENISITHNTRGSESRVALATG	649
QY	1871	STPTACGGCATGATCTACAAACSSCCSACGCTTGACAGCCAGGAAGAAGCACCCSACCTC	1930
Db	650	LEUTIGLIMETILETYLVSPTONISALALENGLINPROGLIYULYSUHNISPROTHVAL	669
QY	1931	CTCTTTGTATATGGAGGCCCCSAGGTSCACSTGGTGAMTATCTCTTCAAGGATCAAG	1990
Db	670	LEUPHEVALITYGLIGLIPROGLINVALGILHEUVALASMSASERHEUSGLIIELYS	689
QY	1991	TACTSGGGGTCACACACACSTGGCTCCCTGGGCTACGGCCGCTGGTGTGATTTGACGGCAG	2050
Db	690	TYTLEUVALGYLEUASNPTRHEUVALASERLENGIYTRALAYALVALIILLESRLYATG	709
QY	2051	GGCTCTCTCAGCAGGGGCTTCGGTTTGAAGGGGCCCCGTAAGAAACCAATGGGCCAGTG	2110
Db	710	GLYSERCYSGLINATRGSLYLEUATGRHEGLIGLIALHEULYASNGINMETGLIGLINVAL	729
QY	2111	GAGATCGAGACACAGCTGGAGGCCCCCACTTGCTGGCCGACAAAGTATGGCTTCATCGAC	2170
Db	730	GLIITGLIUSARGINVALGILUGLILEUGLINRHEVALALAEGLIUSYTGSLYRHEITLASR	749
QY	2171	CTGACCGGAGTTTGCATCCATCGGCTGGCTTACGGGGGGCTCTCTCGTCMATTGGGGCTA	2230
Db	750	LEUSERVALGVALALILEHNISGLYTPRSETTYGLYGLYRHEUHSERLEUHEITGLYLEU	769
QY	2231	ATCCSACAGCCCCSAGTGTTTCAGAGTGCCATCGCAGCGGGTCCCCGGTSCACGCTGGATG	2290
Db	770	ILENLSYSPROGLINVALRHEUVALALALALEAGLYALATGVALIHTVALITRPMET	789
QY	2291	GGCTACGACACAGGGLTACSTGAGCGCTACATGACGCTCCCTGACAAACACACACGCGC	2350
Db	790	ALATYASRPTRHGILYUTYTHRGUATGTUMELASRYVALPROGLIUSASNGINHSIGLY	809
QY	2351	TATGAGGGGGGTTCCGTCGGCTCCSACGTGAGACACSTGCCAAATGAGGCCAAACGGCTTG	2410
Db	810	TYTGLIALAGLISERVALALALEHNISVALGILYULYEURPTOASNGIILPTOASNPARGLEU	829
QY	2411	CTTATCTCCSACAGGCTTCTGTGACGAAGAAGTCCACTTTTCCASACAAATCTCTCTGC	2470
Db	830	LEUILEUHNISGLYRHEUHSERGLIUSNPVALHNISRHEHEHNISITHNINRHEUVAL	849
QY	2471	TCCSACATGATCGACAGGAGAACTTACACGCTCCAGATCTTACSSCAAGACACAC	2530
Db	850	SERGINLEUILEATGALAGLYLVSPTGTYGLILENGLINILETYGPTOASNGIILATGHS	869
QY	2531	AGTATGCTCCGCCCCGAGTGGGGGAGACACTAAGATGACACSTGCTGTGACATTTCTACAG	2590
Db	870	SERILEATGCSPTGGLUSERGILGHNISITRGILVALIHTHEUHNISRHEUGLIN	889
QY	2591	GAATACSTC 2599	
Db	890	GLUTYRILEU 892	
RESULT 4			
AAE24168	ID	AAE24168 standard; Protein; 969 AA.	
XX	AC	AAE24168;	
XX	DT	23-SEP-2002 (first entry)	
XX	DE	Human dipeptidyl peptidase 9 (DPP9) protein.	
XX	KM	Human; dipeptidyl peptidase; DPP; neoplasia; type II diabetes; cirrhosis	
XX	KM	autoimmunity; human immuno deficiency virus; HIV infection; cytostatic;	
KM		graft rejection; antidiabetic; antiinflammatory; immunosuppressive;	
KM		antiviral; enzyme.	

Db 587 CysProIleuysgluIleAlaLeuThrserygluIuTpgluValLeuAlaArgHis 606
 QY 1511 GGTCTCAAGATGTGGTCAATAGAGACAGCAAGCTGTGTACTTCCAGGCAAGAC 1570
 |||||||
 Db 607 GlySerIuIleTrrpValAsnGluGluThrIuysLeuValTyrpneGluGlyThrIuysasp 626
 QY 1571 AGCGCGGTGAGACACACCTCTACGTGTGACATATGAGGCGCGCGGAGATCGTACGC 1630
 |||||||
 Db 627 ThrProIeuGluHisIleuTyrValValSerTyrGluAlaIaGluIleValArg 646
 QY 1631 CTCACACAGCGCGCTTCCATAGCTGCCTCCATGAGCCAGACTCTGACATGTGCTC 1690
 |||||||
 Db 647 LeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheVal 666
 QY 1691 AGCCACTACACAGCGGTGAGACAGCGCGCTGTGACAGCTCTACAGCTACAGCGCGCC 1750
 |||||||
 Db 667 SerHisTyrSerSerValSerThrProProCysValHisValTyrIuysLeuSerGlyPro 686
 QY 1751 GAGACAGACCGCGCTGACACAGCGCGCTGTGCGCTACATGATGAGAGCGACCGCAGC 1810
 |||||||
 Db 687 AspAspAspProLeuHisIuysGluProArgPheTrpAlaSerMetGluAlaAlaSer 706
 QY 1811 TGCCCCCGGATTATGTTCCCTCCAGAGATCTTCATTTCCACAGCGCTCGGATGGCGG 1870
 |||||||
 Db 707 CysProProAspTyrValProProGluIlePheHisPheHisTrpArgSerAspValArg 726
 QY 1871 CTCTACGCGATGATCTACAGACCGCGCTGTGACAGCGAGGAGAAAGCACCGCCAGCTC 1930
 |||||||
 Db 727 LeuTyrGluMetIleTyrIuysProHisAlaLeuGlnProGlyIuysHisProThrVal 746
 QY 1931 CTCTTTGTATATGAGCGCGCGCGAGTGCAGCTGTGAATACTCTTCAAGGATCAAG 1990
 |||||||
 Db 747 LeuPheValTyrGlyGlyProGluValGluLeuValAsnAsnSerPheIuysGlyIleIuys 766
 QY 1991 TACTTGGCGCTCACACACCTGCGCTCCCTGGGCTACCGCGCTGTGTGATGAGCGGACG 2050
 |||||||
 Db 767 TyrLeuAspTyrLeuAsnThrLeuAlaSerLeuGlyTyrValValValIleAspGlyArg 786
 QY 2051 GCGTCTGTCTAGCGAGGCGCTTCGTCAGAGGCGCGCTGAAAAACCAATAGGCGCAGGTG 2110
 |||||||
 Db 787 GlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuIuysAsnGluMetGlyGluVal 806
 QY 2111 GAGATCGAGGACCAAGGTGAGGCGCGCTGACATTCGTGCGCGGAAAGTATGCTTACATGAC 2170
 |||||||
 Db 807 GluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluIuysTyrGlyPheLeuasp 826
 QY 2171 CTGACCGAGTGGCATCATGCGTGTCTACAGGCGCGCTCTCTGCTCATAGGCGGCTA 2230
 |||||||
 Db 827 LeuSerArgValAlaIleHisGlyTrrpSerTyrGlyIuysPheLeuSerLeuMetGlyLeu 846
 QY 2231 ATCCACAAGCGCCAGGTGTCAGAGTGGCCATCGCGGCGTCCCGCTCACCGCTGTGATG 2290
 |||||||
 Db 847 IleHisIuysProGluIuysValPheIuysValAlaIleAlaGlyAlaProValThrValTrrpMet 866
 QY 2291 GCGTACGACACAGGCTACACTGAGCGCTACATGAGCGTCCCTGAGAACACACAGCAGCGC 2350
 |||||||
 Db 867 AlaTyrAspTrpGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGluHisGly 886
 QY 2351 TATAGAGGCGGTCCGCGCGCTGACAGTGGAGAAAGTGGCCATGAGCGCGCAACCGGCTG 2410
 |||||||
 Db 887 TyrGluAlaGlySerValAlaLeuHisValGluIuysLeuProAsnGluProAsnArgLeu 906
 QY 2411 CTATTCCTCCAGCGCTCTCTGAGCAAAACGTGCACCTTTTCCACACAACTCTCTGCTC 2470
 |||||||
 Db 907 LeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHisTrpAsnPheLeuVal 926
 QY 2471 TCCCACTGATCGAGCAGGAAACCTTACAGCTCCAGATCTACCCCAAGACAGACAC 2530
 |||||||
 Db 927 SerGlnLeuIleArgAlaGlyIuysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 946
 QY 2531 AGTATTGCGTGGCGCGGTGGGAGAGACATGAGTGAAGCGAGTGGCTCCACTTCTACAG 2590
 |||||||
 Db 947 SerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGln 966

QY 2591 GAATACCTC 2599
 |||||||
 Db 967 GluTrrpLeu 969
 RESULT 5
 ID ABG61607 standard; Protein: 879 AA.
 XX
 AC ABG61607;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE Human DPPR-2 splice variant #6.
 DE
 KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPR;
 KW DPPRIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
 KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
 KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
 KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
 KW dyskinesia; reproductive disorder; inflammatory disorder;
 KW metabolic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200231134-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US31874.
 XX
 PR 12-OCT-2000; 2000US-240117P.
 XX
 PA (FERR) FERRING BV.
 XX
 PI Qi S, Akinsanya KO, Riviere PJ, Junien J;
 XX
 DR WPI: 2002-444178/47.
 DR N-PSDB: ABK83338.
 XX
 PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
 PT the proteins; useful for treating e.g. fungal, bacterial, protozoan and
 PT viral infections, cancers, allergies, neurological disorders, or pain
 PT -
 PS
 XX
 PS Disclosure; Page 91-93; 113pp; English.
 CC The present invention relates to the isolation of novel human serine
 CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
 CC proteins (DPPR). The dipeptidyl peptidase IV-related proteins (DPPR)
 CC and nucleic acids encoding them are useful for treating infections
 CC such as fungal, bacterial, protozoan and viral infections, particularly
 CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
 CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
 CC bulimia, Parkinson's disease, acute heart failure, hypotension,
 CC hypertension, urinary retention, osteoporosis, angina pectoris,
 CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
 CC psychotic and neurological disorders (e.g. anxiety, dementia, or
 CC schizophrenia), and dyskinesias. These may also be used in discovering
 CC therapeutic agents for the treatment of reproductive, inflammatory and
 CC metabolic disorders. ABG61591-ABG61612 represent human DPPR proteins.
 CC
 XX
 SQ Sequence 879 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 879
 Score: 4558.50 Matches: 850
 Percent Similarity: 98.49% Conservative: 0
 Best Local Similarity: 95.49% Mismatches: 0
 Query Match: 95.07% Indels: 13
 DB: 23 Gaps: 1

US-09-976-674-4 (1-2617) x ABG61607 (1-879)

OY	11	ATGCGCCACACAGGGAGCCCAACGCGCCAGACGAGCGACGCGCCCAACAGATGACCGC	70
Db	30	MetaIaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaAlaIaThrAspAspPro	49
OY	71	GGCGCGCGGCTTCAGGGGCGAGGACGACGCTGTGGAGCGGGGCTCGGAGATCATCCAGCGC	130
Db	50	AlaIaIaArgPheGlnValGlnLysHisSerThrPaspGlyLeuArgSerTleIleHisGly	69
OY	131	AGCGCGAAGACTCGGGCGCTCATTTGTCACAAAGGCGCCCAACAGATTCAGATTGTGCAG	190
Db	70	SerArgLysTyrSerGlyLeuIleValanLysAlaProHisAspPheGlnPheValGln	89
OY	191	AAGACGATAGTCTGGGCGCCCACTCCCAACGCGCTCTACTACCTGGGAATGCCATATGCC	250
Db	90	LysThrAspGlnSerGlyProHisSerHisArgLeuTyrTyrLeuGlnYmeProGlyGly	109
OY	251	AGCGGAGAGAACTCCCTCTACTCTGTAGATTTCCAGAAAGTCCGGAAAGAGGCTGTG	310
Db	110	SerArgLysnSerLeuLeuTyrSerGlnIleProLysValArgLysGlnAlaLeu	129
OY	311	CTGCTCTGTCCTGAGAGCAGATGCTGATCATTTCCAGGCGACCGCCACCATGGGGCTC	370
Db	130	LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnIaIaThrProHisIleGlyAla	149
OY	371	TACTCTCGGGAGAGAGAGCTGTCTAGAGGAGCGGAAAGCGCTGGGGGCTTCGGCATCAC	430
Db	150	TyrSerArgLysGlnGlnIuleuLeuArgGlnArgLysArgLeuGlnYValPheGlyIleThr	169
OY	431	TTCCATCAGACTTCCACAGCGAGATGGGCTTCTCTCTTCCTCCAGGCCAGACAGCCTTTC	490
Db	170	SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnIaIaSerAsnSerLeuPhe	189
OY	491	CACGTGTGCGCGCGCGCGCAGAAAGCGCTCATGTGTCTCCCTATGAAACCGGTGGAATTC	550
Db	190	HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIule	209
OY	551	AAGACCCAGTGTCTAGGGCGCCCGGATGGAACCCCAAAATGTGGCCTGCGACCTGCTTC	610
Db	210	LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProIaIaPhe	229
OY	611	TTTCTCTTCATCAATTAACAGCGACTGTGGGTGGCCACATGAGACAGCGAGACGG	670
Db	230	PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnIleThrGlyGlnIuArg	249
OY	671	CGGTCAGCCTTCTCCCAACAAAGGTTATCCAAATGCTCGGATGAGCCCAAGCTCGCGGT	730
Db	250	ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerIaGly	269
OY	731	GTCGCACTTCGTCATACAGAGAGATTTCGACCGCTTCACGCGGACTGCGTGGCGCCC	790
Db	270	ValaIaThrPheValIleGlnGlnIuPheAspArgPheThrGlyTyrTrpTrpCysPro	289
OY	791	ACAGCCTCTCGGAAGGTTTGAAGGGCTTCAGACGCTGCGAATCCTGTATGAGGAAGTC	850
Db	290	ThrIaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnIuVal	309
OY	851	GATAGTCGCCAGGGGAGAGGATTCATCAAGTCCCTCTCCGCGCTAGAAAGAAAGAACG	910
Db	310	AspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGlnIuLysTyrThr	329
OY	911	GACTCTGATCGGTCACCCAGAGAGCGAGCAAAATCCCAATTTGCTCGAAATCGCT	970
Db	330	AspSerTyrArgTyrProAlaGlnThrGlnGlySerLysAsnProLysIleAlaLeuLysLeuAla	349
OY	971	GAGTTCCAGACTGACAGCCAGGGCGAAGATGTTCTGACCCAGAGAGAAGAGCTGGTCAG	1030
Db	350	GlnPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnIuGlnLeuValGln	369
OY	1031	CCCTTCAGCTGCGCTTCCCGGAAGGTGAGATACATGCGCAGGCGCGGGGTGACCCGGAT	1090
Db	370	ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTyrPheArgAsp	389

QY	1091	GGCAATATACGCTTGGGCAATGTTCTTGACAGCGGCCACAGTGGCTCCAGCTGCTCTC	1150
Db	390	GIlySTyTATATATpAlaMeTpeHeuAspATgProGInGIntPTpLeuGInLeuValLeu	409
QY	1151	CTCCCGCGGGCCGTTCACTCCCGACACAGAAATGAGAGACACGGGCTAGCCCTAGCC	1210
Db	410	LeuProPAlaLeuPheLeuPheTLeuProSerThnGInAsnGInGInArgLeuAlaSerAla	429
QY	1211	AGAGCTGTCCCAAGAAATGTCCAGCCGTTATGTGTCTACAGAGAGTCCACAAAGTCTAG	1270
Db	430	ArgAlaValProArgAsnValGInProTyValValTyTGInGInValThnAsnValTyr	449
QY	1271	ATCAATGTTATGACATCTTCTATCCCTTCCCAATCAGAGGAGAGAGACAGCTGTGC	1330
Db	450	IleAsnValHisAspTLeuPheTyrProPheProGInSerGInGInGInValAspGInLeuGIn	469
QY	1331	TTTTCCGGCGCAATGAATGCAAGACCGGCTTGTGCATTTGTACAAATGCACCGGCTT	1390
Db	470	PheLeuArgAlaAsnGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn	489
QY	1391	TTAAATCCCAAGGCTACGATTTGGATGTGAGCCCTTGACGCCCGGAGAAATGAATTTAA	1450
Db	490	LeuysSerGInGInTyTAspTyrPserGInProPheSerProGInGInAspGInPheGIn	509
QY	1451	TGCCCATTTAAGAAGAGATGTGCTGTACACAGCGGTGAATGGAGAGTTTGGAGAGCAC	1510
Db	510	CysProTLeuGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn	529
QY	1511	GGCTCCCAAGATCTGGGTCAATGAGAGAGCAACAGCTGTACTTCCAGAGGCAACAGGAC	1570
Db	530	GIlySTyTys-----GIlyThnTysAsp	536
QY	1571	ACGGCGGTGAGACACACACTGTACGTGTGCATATGAGAGCGCGCGGAGATCGTACGC	1630
Db	537	ThnProLeuGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn	556
QY	1631	CTCACACAGCGCGGCTTCTCCATAGCTGTCTCATAGGACAGAACTTGACATGTTGCTC	1690
Db	557	LeuThnThnProGInPheSerHisSerCysSerMetSerGInAsnPheAspPheVal	576
QY	1691	AGCCACTACACACCGAGACAGCGCGGCTGGGTGCACGTGTACAAAGTGAAGCGGCCCC	1750
Db	577	SerHisTyrSerSerValSerThnProProCysValHisValTyGInLeuSerGInPro	596
QY	1751	GACGACGACCCCTCGACAAAGACAGCGCGGCTTGGGCTAGCATGATGAGAGCACCGAC	1810
Db	597	AspAspAspProLeuHisTyrGInProArgPheTyrAlaSerMetMetGInAlaAlaSer	616
QY	1811	TGCCCGCCGATATATGTTCTCTCCAGAGATCTTCATTTCCACACCGGCTCGATGTGCGG	1870
Db	617	CysProProAspTyTValProProGInGInPheHisPheHisThnArgSerAspAlaArg	636
QY	1871	CTCTACGGCATGATCTACAAAGCCCGACGCTTGCACACAGGAAAGACACCCACGCTC	1930
Db	637	LeuTyTLeuMetTLeuTyTysProHisAlaLeuGInProGInTyGInGInHisProThnVal	656
QY	1931	CTCTTTGATATGAGAGCGCCCGACAGTCCAGCTGGTAATACCTCTCAAAAGCATTCAG	1990
Db	657	LeuPheValTyTyrGInGInProGInValGInLeuValAlaAsnSerPheLeuGInTyGIn	676
QY	1991	TACTTGGCGCTCAACACACTGGCTCCGTGGCTACAGCGCGTGGTGTGATTCAGCGCAGG	2050
Db	677	TyrLeuArgLeuAsnThnLeuAlaSerLeuGInTyTAlaValValValAlaLeuAspGInTyG	696
QY	2051	GGCTCTCTGACGACGAGGCTTCGGTTTGAAAGGGGCCCTGAAAACCAATGGGCGAGTG	2110
Db	697	GIlySTyCysGInAsnGInTyLeuArgPheGInGInGInGInGInGInGInGInGInGInGIn	716
QY	2111	GAGATCGAGACACAGGTGGAGGGGCTGCATGTGTCGGCCAGAGATGAGGCTTCATGCAC	2170
Db	717	GIuIleGInAspGInValGInGInGInLeuGInPheValAlaGInGInTyTLeuTyTLeuPhe	736
QY	2171	CTGACCCAGTTGCCATTCATAGCTGTGTCTCTACAGGGGGGCTTCTCTGTGCTCATGGGCGCTA	2230

```
Db 737 LeuSerArgValAlaIleHisGlyTyrPheSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 756
OY 2231 ATCCACAGCCCGGAGGTCGTCAGAGTGGCGATGCGGGTGCCCGGTCACCGTGGATG 2230
Db 757 IleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValIthrValTyrMet 776
OY 2291 GCCTACGACACAGGCTACACTGAGCGCTACATGGACGCTCCCTGAGAACACAGCAGCGC 2350
Db 777 AlaTyrAspThrGlyTyrIthrGlnArgTyrMetAspValProGlnAsnGlnHisGly 796
OY 2351 TATGAGCGGGGTTCCGTCGGCCCTGCACAGTGGACAGAGTGGCCCAATGAGCCCAACGCTTG 2410
Db 797 TygGlnAlaGlySerValAlaLeuHisValGlyLysLeuProAsnGluProAsnArgLeu 816
OY 2411 CTATATCCCTCAGCGCTCTCGAGCAAGAAAGCTGATTTTCCACAAATTCCTGTC 2470
Db 817 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 836
OY 2471 TCCCACTGATCGAGCAGGAGAAACCTTACACAGCTCCAGATCTACCCCAACGAGACAC 2530
Db 837 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgHis 856
OY 2531 AGATTCGCTGCCCCGAGTGGCGGAGCAGCTATGAACTCACGTTGCTGCACTTTCTACAG 2590
Db 857 SerIleArgCysProGlnSerGlyGlnHisTyrGlyValIthrLeuLeuHisPheLeuGln 876
OY 2591 GAATACCTC 2599
Db 877 GluTyrIleu 879

RESULT 6
ABG61608
ID ABG61608 standard; Protein; 879 AA.
XX
AC ABG61608;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #7.
XX
KM Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KM DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KM diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KM heart failure; hypertension; urinary retention; osteoporosis; cancer;
KM ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KM dyskinnesia; reproductive disorder; inflammatory disorder;
KM metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI Q1 S, Akinsanya KO, Riviere PJ, Junten J;
XX
DR MPI; 2002-444178/47.
XX
DR N-PSDB; ABR83339.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure; Page 95-96; 113pp; English.
XX
```

```
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bullimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinasias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 879 AA:
Alignment Scores:
Pred. No.: 0 Length: 879
Score: 4558.50 Matches: 850
Percent Similarity: 98.49% Conservative: 0
Best Local Similarity: 98.49% Mismatches: 0
Query Match: 95.07% Indels: 13
DB: 23 Gaps: 1

US-09-976-674-4 (1-2617) x ABG61608 (1-879)
OY 11 ATGGCCACACCGGGAGCCGACGCGGCGAGGCGAGCGAGCGCGGACGATGACCG 70
Db 30 MetAlaThrThrGlyIthrProThrAlaAspArgGlyAspAlaAlaIthrAspAspPro 49
OY 71 GCCGCCCGCTTCCAGGTGCAGAGCACTCTGGAGCGGCTCCGAGCATCATCACGCG 130
Db 50 AlaAlaArgPheGlnValGlnLysHisSerTyrAspGlyLeuArgSerIleIleHisGly 69
OY 131 AGCCGCAAGTACTCGGGCTTCATTTGTCACAGAGGCGCCCAAGACTTCCAGTTTGTCAG 190
Db 70 SerArgLysTyrSerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGln 89
OY 191 AAGACGATGAGTCTGGGGCCCACTCCACCGGCTCTACTACCGGGAGAAATGCCATATGCG 250
Db 90 LysThrAspLysSerGlyProHisSerHisArgLeuTyrIleuGlyMetProTyrGly 109
OY 251 AGCCGAGAGACTCCCTCTCTACTGTGAGATTCCCAAGAGTCCGGAAGAGCTCTG 310
Db 110 SerArgGlnAsnSerLeuLeuTyrSerGluIleProLysValArgLysGluAlaLeu 129
OY 311 CTGCTCTGTCTCGGAGGAGATGCTGTGATCATTTCCAGGCGCACGCCCAACATGGGCTC 370
Db 130 LeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
OY 371 TACTCTCGGAGAGGAGAGCTGTAGAGGAGGAGAAAGCTGGGGGTCTTCGCGATCAC 430
Db 150 TyrSerArgGluGluGlnLeuLeuArgGlnArgLysArgLeuGlyValPheGlyIleThr 169
OY 431 TCCTACGACTTCCACAGCAGAGTGGCTTCTCTTCAGAGGCGACCAAGAGCTCTTC 490
Db 170 SerTyrAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
OY 491 CACTGTGCGCAGCGGCGGCAAGAGGCTTCATGTGTCCCTATGAACCGCTTGAATC 550
Db 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIle 209
OY 551 AAGACCCAGTGTCCAGGCGCCGCGATGAGCCCAAAATTCCTCGCGAGCCGCTTC 610
Db 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
OY 611 TTCTCTTCATCATTAACAGCAGCTGTGGTGCGCAACATCGACAGCGAGGAGCGG 670
Db 230 PheSerPheIleAsnAsnSerAspLeuTyrValAlaAsnIleGluThrGlyGluGluArg 249
OY 671 CGGCTGACCTTTCGCCACCAAGGTTTATCCAAATGCTCTGGATGACCCCAAGTCTGCGGCT 730
|||||
```

Db 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspProLysSerAlaGly 269
 QY 731 GTGGCAACCTTCGTCATPACAGAGAGTTGACCGCTTCACCTGGGTACTGGTGGGCC 790
 Db 270 ValAlaThrPheValIleGlnGlnGlnPheAspArgPheThrGlyTyrTrpPro 289
 QY 791 ACAGCCTCTGGAGAGTTTCAGAGGGCTCAAGAGCTGCAGAACTCTGATAGAGAGTC 850
 Db 290 ThrIleSerTrpGlnGlySerGlnGlyLeuSerThrIleAsnGlyIleLeuTyrGlnGlnVal 309
 QY 851 GATAGATCCGAGGTGGAGGTATTCAGTCCCTCTCTCTGGCTAGAGAAAGAGACG 910
 Db 310 AspIleSerGlnValIleGlnValIleHisValProSerProAlaLeuGlnGlnArgTyrThr 329
 QY 911 GACATGATCGGTACCCAGAGACGGCAGAGAAATCCCAAGATTGCCCTGAACCTGGCT 970
 Db 330 AsperTyrAlaTyrProAlaThrGlnGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
 QY 971 GAGTTCAGACTGACAGCCAGAGGAGATCGTCTGACCCAGAGAGAGAGAGAGCTGGTCAG 1030
 Db 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnGlnLeuValGln 369
 QY 1031 CCCTTCAGCTCGCTGTTCCCAAGGTGAGTACATGCCAGAGGCCGGGTGACCCGGGAT 1090
 Db 370 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTyrTrpArgAsp 389
 QY 1091 GGCAAAATACGCTGGCCATCTTCCTGACCGGCCCCAGCATGGCTCCACTGCTCTC 1150
 Db 390 G1yLysTyrAlaTrpAlaMetPheLeuAsnArgProGlnGlnTrpLeuGlnLeuValLeu 409
 QY 1151 CTCCCGCCGCGCTGTTTCATCCGACAGACAGAAATGAGAGAGAGAGAGAGCTGCTGCC 1210
 Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnGlnArgLeuAlaSerAla 429
 QY 1211 AGAGCTGTCCCAAGATGTCACCGCTATGTGTGTAAGAGAGAGTACCAGACGTGG 1270
 Db 430 ArgAlaValProArgAsnValGlnProTyrValValTyrGlnGlnValThrAsnValTrp 449
 QY 1271 ATCATGATGTCATGACATCTTATCCCTTCCTCCCAATCAGAGAGAGAGAGAGAGCTGTCGC 1330
 Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlnGlnLysAspGlnLeuLys 469
 QY 1331 TTCTCCGCGCCCAATGATGACAGACCGGCTTGCACATTTGTACAAAGTACACCGCTT 1390
 Db 470 PheLeuArgAlaAsnGlnGlySerThrGlyPheCysHisLeuTyrLysValThrAlaVal 489
 QY 1391 TTAATAATCCAGGCTACGATTTGAGTAGACCTTCACCGCCCGGGAAGATGAATTAAAG 1450
 Db 490 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlnLysAspGlnPheLys 509
 QY 1451 TGCCCATTAAGAAAGATGCTCTGACAGCGGTGAATGGAGGTTTGGCGAGGAC 1510
 Db 510 CysProIleLysGlnGlnIleAlaLeuThrSerGlnGlnTyrGlnValLeuAlaTrpHis 529
 QY 1511 GGCTCCAGAGATCTGGGTCAATGAGAGACCAAGCTGTGTACTTCCAGGACCAAGAGAC 1570
 Db 530 GlySerLys-----GlyThrLysAsp 536
 QY 1571 ACGCGCGTGGAGACACACCTTACGTGTCACATAGAGGGCGCGGAGAGATCTGACGC 1630
 Db 537 ThrProLeuGlnHisIleLeuTyrValValSerTyrGlnAlaIleAlaGlnIleValArg 556
 QY 1631 CTCACGACGCGCGGTCTCCCATAGCTGTCACATGACCGCAAGATTCGACATGTTGCTGC 1690
 Db 557 LeuThrThrProGlyPheSerHisSerLysSerMetSerGlnAsnPheAspPheVal 576
 QY 1691 AGCCACTACAGAGCTGAGACAGCGCGCTTCGTCACAGCTTCAAGCTGACGCGGCC 1750
 Db 577 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 596
 QY 1751 GACGACGACCCCTGCACACAGAGCGCGCTTTCGGCTAGCATGATGAGAGACCGACG 1810
 Db 597 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGlnAlaIleAlaSer 616

QY 1811 TGCCCCCGAGATTATGTTCTCTCAGAGATCTTCATTTSCAACAGCGCTCGGATGGCG 1870
 Db 617 CysProProAspTyrValProProGlnIlePheHisPheHisIleArgSerAspValArg 636
 QY 1871 CTCACGCGCATGATCTACAAAGCCCCACGCTTCGACCGCGGAGAGAAACACCCACCGCTC 1930
 Db 637 LeuTyrGlnMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 656
 QY 1931 CTCTTTGATATGAGGCCCCCAGGTGACGTGGTGAATTAATCTCTTAAAGCATCAAG 1990
 Db 657 LeuPheValTyrGlnGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 676
 QY 1991 TACTTGGCGCTCAACACACCTGCGCTCCCTGGGCTAACCGCGGTTCATTTAGAGGCGAG 2050
 Db 677 TyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspValArg 696
 QY 2051 GGCTCTGTTCAGCGAGGCTTCGCTTCGAAAGGGGCCCTGAATAACCAATGGCCAGGCTG 2110
 Db 697 GlySerCysGlnArgGlyLeuArgPheGlnGlnGlnAlaLeuLysAsnGlnMetGlyGlnVal 716
 QY 2111 GAGATCGAGACACAGGTGGAGGCGCTGCAATTCGTGCGCGAAGATAGCTTCATCGAC 2170
 Db 717 GluIleGlnAspGlnValGlnGlnGlnPheValAlaGlnGlnLysTyrGlyPheIleAsp 736
 QY 2171 CTGAGCGGAGTTCGCAATGCGCTGCGTTCACGGGGGCTTCCTGCTCATGGGGCTA 2230
 Db 737 LeuSerArgValAlaIleHisGlyTrpSerTyrGlnGlyPheLeuSerLeuMetGlyLeu 756
 QY 2231 ATCCACAAGCCCCAGGTGTCAGGTGCGCATCGCGGGTGGCCCGCTGACCTGCGATG 2290
 Db 757 IleHisLysProGlnValPheLysValAlaIleAlaGlnGlyAlaProValThrValTrpMet 776
 QY 2291 GCCTACGACACAGGTTACATGAGCGTACATGACAGTCCCTGAGAACACACGACGCGC 2350
 Db 777 AlaTyrAspTrpGlnGlyTyrThrGlnArgTyrMetAspValProGlnAsnAsnGlnHisGly 796
 QY 2351 TATGAGGGGGGTTCGCTGCGCTGACGAGGAGAGAGTGCACCAAGACCCCAACGCGCTG 2410
 Db 797 TyrGlnAlaGlnGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeu 816
 QY 2411 CTATCTCTCAGCGCTCTCTGACGAAAGCGTCTTTTSCAACAACTCTCTGCTC 2470
 Db 817 LeuIleLeuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuVal 836
 QY 2471 TCCCACTGATCCGACAGGGAACCTTACACAGCTCCAGATCTACCCACAGAGACAC 2530
 Db 837 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 856
 QY 2531 AGTATCGCTGCGCGAGTCCGGGCGAGACATGAAAGTCAGCTTGTGCACTTCTACAG 2590
 Db 857 SerIleArgCysProGlnSerIleGlnHisTyrGlnValThrLeuLeuHisPheLeuGln 876
 QY 2591 GAATACCTC 2599
 Db 877 GluTyrLeu 879
 RESULT 7
 AA#24171
 ID AA#24171 standard; protein: 830 AA.
 AC AA#24171;
 DE 23-SEP-2002 (first entry)
 DT
 XX Human dipeptidyl peptidase 4 (DPP4)-like 2 protein.
 XX Human dipeptidyl peptidase: DPP: neoplasia: type II diabetes: cirrhosis;
 KW autoimmunity: human immuno deficiency virus: HIV infection: cytostatic;
 KW graft rejection: antidiabetic: antinflammatory: immunosuppressive;
 KW antiviral: enzyme: DPP-4 like 2 protein.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 235
 FT /note- "Encoded by GAG"
 XX
 XX MO200234900-A1.
 XX
 XX 02-MAY-2002.
 XX
 XX 29-OCT-2001; 2001MO-AU01388.
 XX
 XX 27-OCT-2000; 2000AU-0001078.
 XX
 XX (UNSY) UNIV SYDNEY.
 XX
 XX Abbott CA, Gorrell MD;
 XX
 XX WPI; 2002-454646/48.
 XX
 XX N-PSDB; AAD38957.
 XX
 XX New dipeptidyl peptidase (DPP) peptides, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -
 XX
 XX Disclosure; Page 82-86; 91pp; English.
 XX
 XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polynucleotides encoding such proteins. The DPP peptides are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is human DPP4-like 2 protein.
 XX
 XX
 SO Sequence 830 AA:
 Alignment Scores:
 Pred. No.: 0 Length: 830
 Score: 4458.00 Matches: 828
 Percent Similarity: 99.88% Conservative: 1
 Best Local Similarity: 99.76% Mismatches: 1
 Query Match: 92.97% Indels: 0
 DB: 23 Gaps: 0
 US-09-976-674-4 (1-2617) x AAE24171 (1-830)
 QY 110 CTCGCGAGCATATCCAGCGAGCGGCAAGACTCGGCGCTATTTCACAAAGGCGCC 169
 DB 1 LeuArgSerIleIleHisGlySerArgLysTyrSerGlyLeuIleValAsnLysAlaPro 20
 QY 170 CAGGACTTCCAGTTTGTGCAGAAAGCAGATGAGTGTGGGCCCACTCCACCGCTTAC 229
 DB 21 HisAspPheGlnPheValGlnLysThrAspGlnSerGlyProHisSerHisArgLeuTyr 40
 QY 230 TACCTGGGAATGCCATATGCGAGCGGAGAACTCCCTCTCTACTGTAGATTTCCAG 289
 DB 41 TyrLeuGlyMetProTyrGlySerArgGlnAsnSerLeuLeuTyrSerGlyIleProLys 60
 QY 290 AAGTCGCGAAGAGGCTGTGCTGCTCTGCTCCGGAAGAGAGATGCGATTCATTTCAG 349
 DB 61 LysValArgLysGlnAlaLeuLeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGln 80
 QY 350 GCCAGCGCCACCATGGGCTCTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAAGCGC 409
 DB 81 AlaThrProHisHisGlyValTyrSerArgGlnGlnIleuLeuArgGlnArgLysArg 100
 QY 410 CTGGGGGTCTTGGGATACCTCTTACGACTTCCACAGCGAGAGTGCGCTTCTCTTTC 469
 DB 101 LeuGlyValPheGlyIleThrSerTyrAspPheHisSerGlnSerGlyLeuPheLeuPhe 120
 QY 470 CAGGCGAGCAAGACCTTCCACTGTCCGAGAGCGGCGGCAAGAGGCTTCATGTTGCC 529
 DB 121 GlnAlaSerAsnSerLeuPheHisCysArgAspGlyGlyLysAsnGlnPheMetValSer 140

QY 530 CCTATGAACCGCTGGAATCAAGACCAGTGTCTAGGGCCCCGGATGAGCCCAATC 589
 DB 141 ProMetLysProLeuGlnIleLeuTyrGlnClnSerGlyProArgMetAspProLysIle 160
 QY 590 TGGCCTCGCGACCTGCTTCTCTCTTCATCAATTAACAGCGACTGTGGGTGCCAAC 649
 DB 161 CysProAlaAspProAlaPhePheSerPheAsnAsnAsnSerAspLeuTyrValAlaAsn 180
 QY 650 ATGAGACAGCGCGAGGCGGCGGTGACCTTGTCCACCAAGGTTATTCAGTCTCTG 709
 DB 181 IleGluThrGlyGlnGlnArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 200
 QY 710 GATGACCCCAAGTCTGCGGGTGTGGCCACTTCTGCTATACAGAAAGATTTCAGCCGCTTC 769
 DB 201 AspAspProLysSerIleArgLysAlaIleThrPheValIleGlnGlnGlnIlePheAspArgPhe 220
 QY 770 ACTGGTACTGTGTGTGCTCCACAGCTCTCTGGAGAGTTTCAGAGGCTTCAGAGCTG 829
 DB 221 ThrGlyTyrTrpTrpCysProThrAlaSerTrpGlnGlySerGlnGlyLeuTyrThrLeu 240
 QY 830 CGAATCCTGATGAGGAAGTCGATGATCCGAGGTGAGGTCATTCACGTCCTCTCTCT 889
 DB 241 ArgIleLeuTyrGlnGlnValAspGlnSerGlnValIleValIleHisValProSerPro 260
 QY 890 GCGCTAGAAAGAAAGAGAGAGGAGACTCGTATCGGTACCCAGAGAGAGACCAAGATCCC 949
 DB 261 AlaLeuGlnGlnArgLysThrAspSerTyrArgTyrProArgThrGlySerLysAsnPro 280
 QY 950 AAGATTCCTTGAATGCTGCTGATGCTCCAGACTGACAGCGGCAAGATGCTGTGACC 1009
 DB 281 LysIleAlaLeuLysLeuAlaGlnPheGlnThrAspSerGlnGlyIleValSerThr 300
 QY 1010 CAGGAGAAAGAGTGTGTCAGAGCTGTGACGCTGTCTCCGAAGGTGAGTATATGCC 1069
 DB 301 GlnGlnLysGlnLeuValGlnProPheSerSerLeuPheProLysValGlnTyrIleAla 320
 QY 1070 AGGCGCGGTGAGACCCGCGATGCGAATACGCTGTGGCCATGTTCTGAGCCGCGCCAG 1129
 DB 321 ArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGln 340
 QY 1130 CAGTGGCTCCAGCTGCT 1189
 DB 341 GlnTrpLeuGlnLeuValIleLeuProProAlaLeuPheIleProSerThrGlnAsnGln 360
 QY 1190 GAGCAGCGGCTAGGCTGTGCGAGAGTGTGCGGAGATGCGAGGCTGTGATGTGTGATC 1249
 DB 361 GlnGlnArgLeuAlaSerAlaArgAlaValProArgAsnValGlnProTyrValIleTyr 380
 QY 1250 GAGGAGTCAACAAGCTGTGATCAATGTTATGACATCTTATTCCTTCCCAATCA 1309
 DB 381 GlnGlnValIleHisValIleTrpIleAsnValHisAspIlePheTyrProPheProGlnSer 400
 QY 1310 GAGGAGAGAGACAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1369
 DB 401 GlnGlyGlnLysProLysCysPheLeuArgAlaAsnGlnLysLysTrpGlyPheCysHis 420
 QY 1370 TTGTACAAAGTCAACCGCGCTTTTAAATCCAGAGGCTACGATGGAGTGGAGCCTTAGC 1429
 DB 421 LeuTyrLysValIleHisValIleLeuLysSerGlnGlyTyrAspTrpSerGlnProPheSer 440
 QY 1430 CCGCGGGAAGATGAATTTAAGTCCCATTAAGAGAGATGCTGTGACAGCGGTGAA 1489
 DB 441 ProGlyLysProLysPheLysCysProIleLysGlnIleValLeuThrSerGlyGln 460
 QY 1490 TGGGAGTTTGGCGAGCGAGCGGCTCCAGATGTGGTCTCATAGAGAGCAAGCTGTGG 1549
 DB 461 TrpGlnValLeuAlaArgHisGlySerLysIleTyrValAsnGlnGlnTyrHisVal 480
 QY 1550 TACTTCAGGCGCAAGGACAGCGCGCTGAGCAGACACCACTTCACGGTGCACCTATGAG 1609
 DB 481 TyrPheGlnIleTyrLysAspThrProLeuGlnHisIleLeuTyrValValSerTyrGln 500


```

QY 1610 GCGGCGGCGAGATGATACGCTTCACACAGCCGCGCTTCTCCATAGCTGCTCATGAGC 1669
    |||
Db 501 AlAlaIaGlIuIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSer 520
QY 1670 CAGAACTTCGACATGTTGTAGGCACTACAGACGCTGAGACGCGCGCTGCGTGAC 1729
    |||
Db 521 GlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHis 540
QY 1730 GTCTACAAAGCTGAGCGCGCGGAGAGACGCCCTGCACAGACGCGCGCTTGCGGCT 1789
    |||
Db 541 ValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheThrPala 560
QY 1790 AGCATGATGAGAGCAGCAGCTGCCCGCGGATTATGTTCTCCAGAGATCTTCATTC 1849
    |||
Db 561 SerMetMetGluAlaIaIaSerCysProProAspPyrValProProGluIlePheHisPhe 580
QY 1850 CACACGCGCTCGATGTGCGGCTTACGGCATGATCTACAAAGCCCGACGCTTGACGCA 1909
    |||
Db 581 HisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaIeucIlnPro 600
QY 1910 GGGAAAGAGCAGCCCGCGCTCTTGTATATGAGGCGCGCGAGCTGAGCTGAT 1969
    |||
Db 601 GlyLysLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsn 620
QY 1970 AACTCTTCAAGGCAATCACTACTGCGGCTCAACACACAGCTGCGCTCCCTGCGGTACGCC 2029
    |||
Db 621 AsnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAla 640
QY 2030 GTGTTGTGATTTAGCGGCGGCGCTCTGTACGAGGAGGCTTCCGTTGGAAGGCGCGCTG 2089
    |||
Db 641 ValValValIleAspArgValArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeu 660
QY 2090 AAAAACAAGATGGGCGCGGATGAGATGAGAGACAGGATGGAGGCGCTCACTGCTGCGCC 2149
    |||
Db 661 LysAsnGlnMetGlyGlnValGlnIleGlnAspGlnValGlnGlyLeuGlnPheValAla 680
QY 2150 GAGAAGTATGAGCTTCATGACGCTGAGCGGATGCGCATGCGTGGCTGCTACGCGGCGC 2209
    |||
Db 681 GluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrPserTyrGlyGly 700
QY 2210 TTCTCTGCTCATGGGCTTAATCCACAAGCGCGAGTGTTCAGAGTGGCGCATCGCGGCT 2269
    |||
Db 701 PheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGly 720
QY 2270 GCGCGGCTGACCGCTGATGAGCTACGACACAGGCTACATGAGCGCTACATGAGCGCTC 2329
    |||
Db 721 AlaProValIlnrValIlnrPheAlaIlnrAspThrGlyTyrIlnrGluArgTyrMetAspVal 740
QY 2330 CCTGAGAACACACAGCAGCGCTATGAGCGGCTTCCGCGCGCTGACGCTGAGAGAGCTG 2389
    |||
Db 741 ProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeu 760
QY 2390 CCCAATGAGCCCAACCGCTTCTTATCCCTCCAGCGGCTTCTGAGAGAAAGCTGCACATT 2449
    |||
Db 761 ProAsnGlnProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhe 780
QY 2450 TTCACACAAACTTCCTGCTGCTCCCACTGATCCGAGCAGGAGAACTTACAGCTCCAG 2509
    |||
Db 781 PheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGln 800
QY 2510 ATTCACCCCAAGAGAGACAGATATTGCTGCGCGCGAGTGGGCGAGCAGCTATGAATGC 2569
    |||
Db 801 IleTyrProAsnGlnIlnrHisSerIleArgCysProGlnSerGlyIlnHisTyrGlnVal 820
QY 2570 ACGTGTGCTGCACTTTCTACAGGAATACCTC 2599
    |||
Db 821 ThrLeuLeuHisPheLeuGlnGlyTyrLeu 830

```

RESULT 8
 AAE24169
 ID AAE24169 standard; Protein; 869 AA.
 XX
 AC AAE24169;

```

XX 23-SEP-2002 (first entry)
DE Alternative version of murine dipeptidyl peptidase 9 (DPP9) protein.
XX
XX Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
XX human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
XX type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
XX antiviral; enzyme.
XX Mus sp.
XX WO200234900-A1.
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-AU01388.
XX
XX 27-OCT-2000; 2000AU-0001078.
XX
XX (UNSY ) UNIV SYDNEY.
XX Abbot CA, Gorrell MD;
XX
XX WPI: 2002-454646/48.
XX N-PSDB: AAD38955.
XX
XX New dipeptidyl peptidase (DPP) peptidases, useful for screening
XX inhibitors of DPP catalytic activity, which may be employed to treat
XX e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV infection -
XX
XX Claim 1: Page 70-74; 91pp: English.
XX
XX The present invention relates to dipeptidyl peptidase (DPP) proteins and
XX polynucleotides encoding such proteins. The DPP peptidases are useful for
XX screening inhibitors of DPP catalytic activity. The inhibitors are useful
XX for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
XX rejection and HIV (human immuno deficiency virus) infection. The present
XX sequence is an alternative version of murine DPP9 protein.
XX Note: This sequence is stated to be the same as that shown as
XX SEQ ID NO: 4 in the sequence listing of the specification. However these
XX sequences differ.
XX
XX Sequence 869 AA:
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 869
XX Score: 4279.00 Matches: 792
XX Percent Similarity: 95.81% Conservative: 32
XX Best Local Similarity: 92.09% Mismatches: 36
XX Query Match: 89.24% Indels: 0
XX DB: 23 Gaps: 0
XX
XX US-09-976-674-4 (1-2617) x AAE24169 (1-869)
QY 20 ACCGGAGCCCAAGCGCGCGAGGCGAGCGCGCGCATGACCGCGCGCGCGC 79
    |||
Db 10 SerGlyValSerProValGlnGlnValAlaIaIaGlyAspMetAspThrAlaIaArg 29
QY 80 TTCCAGGTGCAAGAACACATCGTGGGAGCGGCTCCGCGAGCATGACAGGAGCGCGAG 139
    |||
Db 30 PheCysValGlnLysHisSerTyrAspArgLeuArgSerIleIleHisGlySerArgLys 49
QY 140 TACTCGGAGCTCATGTGCAAGAGCGCGCGCGAGCTTCCAGTTGTGCAGAAAGAGAT 199
    |||
Db 50 SerSerGlyLeuIleValSerTyrLysAlaProHisAspPheGlnPheValGlnLysProAsp 69
QY 200 GAGTGTGGGCGCGCGCATCCACCGCGCTTACTACTGCGGAATGCAATGAGAGCGCGAG 259
    |||
Db 70 GluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGlySerArgGln 89
QY 260 AACTGCCCTGCTACTGAGATTTCCCAAGAAAGTCCGGAAGAGGCTGCTGCTGCTG 319

```

```
|||||
Db 90 AsnSerLeuLeuYrSerGluIleProLysLysValAlrGlySgluaIalaLeuLeuLeu 109
320 TCCGGAAGAGATGCGATTCATTTCCAGGCCAGGCCACCATGGGGGTACTTCGG 379
|||||
Db 110 SerTrpLysGlnMetLysAsnHisPheGlnAlaThrProHisLysGlyValTyrSerArg 129
380 GAGGAGAGAGTGTGTAGGAGACGGAAAGCCTGGGGGTCTTCGGCATCACCTCTACGAC 439
130 GluGluLeuLeuArgGluArgLysArgLengLysAlaPheGlyIleThrSerTyrAsp 149
440 TTCACAGCCAGAGTGGCTTCTCTCTTCAGGCCAGCAACACCTTCTTCATGTGCG 499
Db 150 PheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerHisSerLeuPheHisCysArg 169
500 GACGGCCGCAAGACGGCTTCATGTGTCCTGAAACCGCGGAAATCAACAGCCAG 559
170 AspGlyLysLysAsnGlyPheMetValSerProMetLysProLeuGluLysThrGln 189
560 TGCACAGGCCCGGATGGAACCCAAATCTGCCCTGCCAGCCCTGCTCTCTCC 619
190 CysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPhe 209
620 ATCAATPACAGCACTGTGGGTGGCCACATCGAGACAGCGAGAGCGGCGCTGACC 679
210 IleAsnHisSerAspLeuTrpValAlaAsnIleGluThrGlyGluArgLysArgLeuThr 229
680 TTTCGCCACCAAGGTTATCAATGTCTGGATGACCCCAAGTGTGGGGGTGGCCACC 739
230 PheCysHisGlnGlySerAlaGlyValLeuAspAsnProLysSerAlaGlyValAlaThr 249
740 TTGCTACACAGGAAGATTGACCGCTTCACTGGTACTGGTGCGTGGCCACAGCTCC 799
250 PheValIleGlnGluGluPheAspArgPheThrGlyCysTrpTrpCysProThrAlaSer 269
800 TGGGAAGTTCAGAGGGCCTCAGACCGCTGCAATCTCTGATGAGAAAGTTCGATGAGTCC 859
270 TrpGluLysSerGluGlyLeuLysThrLeuArgIleLeuTrpGluGluValAlaSer 289
860 GAGGTGAGGTCATTCACGTCCCTCTCTGCGCTAGAGAAAGAGAGAGAGATTCGAT 919
290 GluValGluValIleHisValProSerProAlaLeuGluGluValGlyLysThrAspSerTyr 309
920 CGGTACCCAGGACAGGACCAAGATCCCAAGATTCCTGAAATCGCTGAGTTCAG 979
310 ArgTyrProArgTrpArgLysSerLysAsnProLysIleAlaLeuLysLeuAlaGluLeuGln 329
980 ACGTACAGCCAGGCAAGATGCTGTGCAACCCAGAGAAAGAGCTGTGTCAGCCTTACG 1039
330 ThrAspHisGlnGlyLysIleValAspSerCysGluLysGluLeuValGlnProPheSer 349
1040 TTCGTTTCCCGAAGGTGGATACATCGCCAGGCGCGGTGGAGCCGGGATGGCAATAC 1099
350 SerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrAlaGlyAspGlyLysTyr 369
1100 GCCTGGGCACTGTTCTGGAACGGGCCAGACAGTGGCTTCAGCTCTCTCTCCCGG 1159
370 AlaTrpAlaMetPheLeuAspArgProGlnGlnArgLengLysValLeuLeuProPro 389
1160 GCCCTGTTCATCCGAGACACAGAGATGAGAGAGAGCGGCTGAGCTTCGCCAGCTGTC 1219
390 AlaLeuPheIleProAlaValAlaGluSerGluAlaGlnArgGlnAlaAlaAlaValAlaVal 409
1220 CCCAGGAATGTCCAGCGGTATGTGTAGCAGAGAGGTTCACCAACGTCTGGATCAATGTT 1279
410 ProLysAsnValGlnProPheValIleTyrGluGluValThrAsnValIleTrpIleAsnVal 429
1280 CATGACATCTTATCTCTCCCTCCCAATCAGAGGAGAGAGACAGCAAGCTCTGCTTCGCG 1339
430 HisAspTrpIlePheHisProPheProGlnAlaGluGlnGlnAspPheCysPheLeuArg 449
1340 GCCAATGATGCAAGACCGGCTTGCCTATTTGTCAAAAGTCAACGGCGTTTAAATCC 1399
|||||

|||||
Db 450 AlaAsnGluCysLysTrpGlyPheCysHisLeuTyrArgValThrValGluLeuLysThr 469
1400 CAGGCTACGATTTGGAGAGACCTTCAGCCCGGAGAGATGAATTAAGTGGCCATT 1459
::: |||||
Db 470 LysAspTrpArgTrpTrpGluProLeuSerProThrGluGluGluPheLysCysProIle 489
1460 AAGGAAGATTGCTCTGACACGGGTGAATGGAGGTTTTGGGAGGACAGGCTCCAG 1519
490 LysGluGluValAlaLeuThrSerGlyGluTrpGluValLeuSerArgHisGlySerLys 509
1520 ATCTGGGTCATGAGAGACCAACCTGTGACTTCCAGGGCCACCAAGACACCCCTG 1579
510 IleTrpValAsnGluGlnTrpLysLeuValTyrPheGlnGlyTrpLysAspThrProLeu 529
1580 GAGCAGCACCTCTACGTGTCAGCTATGAGCGGCGGAGATCTACGCTCAGCAGC 1639
530 GlnHisHisLeuTyrValIleValSerTyrGluSerAlaGlyGluIleValAlaGlyLeuThr 549
1640 CCCGGCTTCCCATACCTGCTCCATGAGCCAGACCAACTTCGACATGTTCTGACGCCATAC 1699
550 LeuGlyPheSerHisSerCysSerMetSerGlnSerPheAspMetPheValSerHisTyr 569
1700 AGCAGCTGAGACAGCCCGCTGCTGCTCAGCTCTACAAAGCTGAGAGGCGCCGAGAGAC 1759
570 SerSerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspAsp 589
1760 CCCCTGCACAAGACCGCCGCTTCTGGGTACATGATGAGGAGGAGCGAGCTGCCCCG 1819
590 ProLeuHisLysGlnProAlaGlyPheTrpAlaSerMetMetGluAlaAlaAsnCysProPro 609
1820 GATTATGTCCTCCAGAGATCTTCATTTCCACAGCGGCTCGATGTCGGCTTACGCC 1879
610 AspTyrValProProGluIlePheHisPheHisThrArgAlaAspValGlnLeuTyrGly 629
1880 ATGATCTACAAGCCCCACGCTTCGACCAAGGAAGAAGACCCACCTCTCTTTGTA 1939
630 MetIleTyrLysProHisThrLeuGlnProGlyArgLysHisProThrValLeuPheVal 649
1940 TATGAGAGCCCCAGGTGACAGTGTGTAATTAACCTTCAAGAGCAACAAGTACTTCGG 1999
650 TyrGlyLysProGlnValGlnLeuValAlaAsnSerPheLysGlyIleLysTyrLeuArg 669
2000 CTCACACACTGGCTCCCTGGGCTACGCGGTGTTGATTCAGGCGAGGGCTCTGT 2059
670 LeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCys 689
2060 CAGCGAGGCTTCGGCTTCGAAGGGCCCTGAACCAACCAATGGCGCCAGTCAGATGAG 2119
690 GlnArgGlyLeuHisPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGln 709
2120 GACCAGGTGAGGGCTGCAAGTTCGTGGCCGAGAAGTATGAGCTTCATCAGACTGAGCCGA 2179
710 AspGlnValGluGlyLeuGlnTyrValAlaGluLysTyrGlyPheIleAspLeuSerArg 729
2180 GTTGCCATCCATGGCTGCTCAGGGGCTTCTCTGCTCATAGGGGCTAATCCAGAG 2239
730 ValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLys 749
2240 CCCAGGTTCACAGGTGGCCATTCGGGGTGGCCCGCTGACACCTGTGATGGCTTACGAC 2299
750 ProGlnValPheLysValAlaIleAlaGlyAlaProValThrValIleTrpMetAlaTyrAsp 769
2300 ACAAGGTACACTGAGCGCTCATGAGCTCCCTGAGAACAACAGCAGCGGCTATGAGCG 2359
770 ThrGlyTyrThrGluArgTyrMetAspValProLysAsnHisGlnGlnTyrGluAla 789
2360 GGTTCGCTGGCCCTGCAGCTGGAGAGACTGCCCAATGAGCCAAACGCTTGTATCTTC 2419
790 GlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgLeuLeuIleLeu 809
2420 CAGCGCTCTCTGAGCAAAACGTCGACTTTTCCACACAACACTCCGCTGCCAGCTG 2479
810 HisGlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeu 829
|||||
```



```
QY 1031 CCTTCAGCTCGCTGTTCCGAGAGGTGAGTACATCCGACGCGGGGTGACCCGGGAT 1090
    |||
Db 370 ProbeserSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgasp 389
QY 1091 GGCAGAAATACGGCTGGGCAATGCTCTGAGACCGGCCAGAGTGGCTCCACCTGCTCTC 1150
    |||
Db 390 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu 409
QY 1151 CTCGCCCGGCGCTGTCATCCGAGACAGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGG 1210
    |||
Db 410 LeuProProAlaLeuPheIleProSerThrGlnAsnGlnGlnAlaGlyLeuAlaSerAla 429
QY 1211 AGAGCTCTCCGAGGAATGTCACCCGTATGCTGTACGAGGAGGTCCACCACTGCTGG 1270
    |||
Db 430 ArgAlaValProArgAsnValGlnProTyrValAlaTyrGlnGlnValThrAsnValTrp 449
QY 1271 ATCAATGTATGACATCTTATCCCTTATCCCTCCCAATCAGAGGAGGAGGAGGAGGAGG 1330
    |||
Db 450 IleAsnValHisAspIlePheTyrProPheProGlnSerGlnGlnAspGluLeuGlnCys 469
QY 1331 TTTCTCCGCGCAATGATGATGACAGACGCGCTTTCACATTTGTACAAAGTACCGCGCTT 1390
    |||
Db 470 PheLeuArgAlaAsnGlnCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 489
QY 1391 TTAAATCCAGGGCTACGATTGGAGTGAGCCCTTCAGCCCGGGGAGATGAATTTAG 1450
    |||
Db 490 LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGlnGlnAspGluPheLys 509
QY 1451 TGGCCCATTAAGAGAGATTTGCTGTACCAAGGCTGAATCGGAGTTTGGCCAGGAC 1510
    |||
Db 510 CysProLysGlnGlnIleAlaLeuThrSerGlnLysTrpGlnValLeuAlaArgHis 529
QY 1511 GGCTCCAAAGTCTGGGTCAATGAGAGAGACCAAGCTGGTACTTCCAGGCGACCAAGAC 1570
    |||
Db 530 GlySerLysIleTrpValAsnGlnGlnIleThrLysLeuValTyrPheGlnGlnTyrThrLysAsp 549
QY 1571 ACGCGCTGGAGACACACCTCTACGTGTGACGTATAGGCGCGCGAGAGTGTACGC 1630
    |||
Db 550 ThrProLeuGlnHisHisLeuTyrValValSerTyrGlnAlaGlnGlnIleValArg 569
QY 1631 CTGACACGCGCGCTTCCATAGTGTGCTGATGAGCCAGAACTTCGATCTGTGTC 1690
    |||
Db 570 LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspPheVal 589
QY 1691 AGGCATACAGACAGCTGTACAGCCGCCCTGCTGTGACAGCTTACAAAGTGTAGCGGCCCC 1750
    |||
Db 590 SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerGlyPro 609
QY 1751 GAGGAGACCCCTGCGCAAGACAGCCCGCTTGGCTGATGATGAGAGCGCCAGC 1810
    |||
Db 610 AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetGlnAlaAlaSer 629
QY 1811 TGGCCCGCGGATTATGTTCTCCAGAGATCTTCATTTCCACACGCGCTGGATGTGCGG 1870
    |||
Db 630 CysProProAspTyrValProProGlnIlePheHisPheHisThrArgSerAspValArg 649
QY 1871 CTCTACGGCATGTATACAAAGCCCGCTGTGACAGCCAGAGAGAGAGAGAGAGAGAGAG 1930
    |||
Db 650 LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlnLysLysHisProThrVal 669
QY 1931 CTCTTTGTATATGAGAGGCGCCAGGTGACAGCTGGTGAATTAAGTCTTCAAGGCATCAG 1990
    |||
Db 670 LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLys 689
QY 1991 TACTTGCGGTCAACACACAGGCGCTGCGGTGAGCGGTGATGATGAGCGAGG 2050
    |||
Db 690 TyrLeuThrGlyLeuAsnThrLeuAlaSerLeuGlyTyrAlaValAlaIleAspGlyArg 709
QY 2051 GGCCTCTGTACAGAGGCGCTTCGTTTGAAGGCGCGCTGAAACCAATAGGCGAGGTG 2110
    |||
Db 710 GlySerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnVal 729
QY 2111 GAGATGAGAGACAGGTGAGGCGCTGCAAGTTCGTGGCGAGAAATGAGCTTCATCGAC 2170
```

```
Db 730 GluIleGlnAspGlnValGlnGlnIleGlnIlePheValAlaGlnLysTyrGlyPheIleAsp 749
QY 2171 CTGAGCCGAGTGTGCCATTCATAGCGTGTCTACAGGGGGCTTCCTGCTCATGGGCTA 2230
    |||
Db 750 LeuSerArgValAlaIleHisGlyTyrPheSerTyrGlyGlyPheLeuSerLeuMetGlyLeu 769
QY 2231 ATCCACAAGCCCGAGGTTCATGAGTGGCCATCGCGGGTGGCCCGGTCACCGTCTGGATG 2290
    |||
Db 770 IleHisLysProGlnValPheLys--Ala----- 778
QY 2291 GCCTAGACACAGGGTACACTGAGCGCTACATGAGCGTCCCTGAGAACACACAGCAGCGC 2350
    |||
Db 778 ----- 778
QY 2351 TATGAGCGGGTTCGCGGCTTCAGCTGAGAGAGCTGCCCAATGAGCCCAACCGCTTG 2410
    |||
Db 779 -----GlnProLeuA 782
QY 2411 CTATTCCTCCAGGCTTCGTGAGCAAGAAAGTGACATTTTCCACAGCAATTCCTGTC 2470
    |||
Db 782 LaryrProProAlaGlnLeuProGlnArgLysArgAlaLeuPheProHisLysLeuProArgL 802
QY 2471 TCCCACTGATCCGAGGAGGAGAACTTACCAAGCTCCAGATCTACCCCAAGAGAGACAC 2530
    |||
Db 802 euProThrAspProSerArgGlnThrLeuProAlaProAspLeuProGlnArgGlnThrG 822
QY 2531 AGTATTCGCTGCCCGCGAGTGGCGGAGCAGCTA 2562
    |||
Db 822 LntyrSerLeuProArgValGlnArgAlaLeu 832

RESULT 10
ABG61606
ID ABG61606 standard; protein; 832 AA.
XX
AC ABG61606;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #5.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPP1; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinnesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200231134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR ) FERRING BV.
XX
PI O1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR WPI: 2002-444178/47.
XX
DR N-PSDB; ABK83337.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure: Page 88-90; 113pp; English.
```



```
QY 1811 TGCCCCCGGATATGTTCTCTCAGAGATCTTCATTTCCACACGCGCTCGATGTGCGG 1870
|||
Db 630 CysProProAspArgValaProProGluIlePheHisPheHisSThrArgSerAspValaArg 649
|||
QY 1871 CTCACGGCATGATCTACACAGCCCCAGCCTTGACAGCCAGGAAGACACCCACCGCTC 1930
|||
Db 650 LeuTyrGluMetIleTyrLysProHisAlaLeuInProGluLysLysHisProThrVal 669
|||
QY 1931 CTCCTTGATATGAGAGGCCCCAGTGCAGCTGGTGAATATCTCTTCAAGGATCAAG 1990
|||
Db 670 LeuPheValIleTyrGluGlyProGluValaGlnLeuValaIleAsnSerPheLysGlyIleLys 689
|||
QY 1991 TACTTGGCGCTCAACACAGCTGGCTCCCTGGGCTACCGCGCTGTTGCATTTAGCGCAGG 2050
|||
Db 690 TyrLeuArgLeuAsnThrLeuAlaSerLeuGluTyrAlaValaIleAspGlyArg 709
|||
QY 2051 GGCCTCTGTCAGAGAGGCTTCGGTTGCAAGGCGCCCTGAAAGAACCAATAGGCGCAGTG 2110
|||
Db 710 GlySerLysGlnArgGlyLeuArgPheGluGluIleLeuLysAsnGlnMetGlyGlnVal 729
|||
QY 2111 GAGATTCAGACAGCAGTGGAGGCGCTGCAGTTCGTGGCCGAGAGTATGCTTCATCGAC 2170
|||
Db 730 GluIleGluAspGlnValaGluGlyLeuGlnPheValaIleGluLysTyrGlyPheIleAsp 749
|||
QY 2171 CTGAGCCGAGTTGGCCATTCATGCGTGGTCTTACGCGGGGCTTCCTGCGCTCATGGGCTA 2230
|||
Db 750 LeuSerArgValaIleIleHisGlyTyrPserTyrGlyGlyPheLeuSerLeuMetGlyLeu 769
|||
QY 2231 ATCCACAGACCCCGAGTGTTCATGAAGTGCCGATCCGCGGGTGCCTGACCGTCTGCATG 2290
|||
Db 770 IleHisLysProGlnValaIlePheLys--Ala----- 778
|||
QY 2291 GCTACAGACACAGGGGTACACTGACGCGCTACATGAGCGTCCCTGAGAACACACAGCAGCGC 2350
|||
Db 778 ----- 778
|||
QY 2351 TATGAGCGGGTTCCTGGCCCTGCACGTCGAGAGAGCTGCCCAATGAGCCCAACCGCTTG 2410
|||
Db 779 -----GlnProLeuA 782
|||
QY 2411 CTATTCCTCCACGGCTTCCTGAGAGAAACGTGCATTTTCCACACAAACTTCTCTGCTC 2470
|||
Db 782 IatyrProProAlaArgLeuProGlyArgLysArgAlaLeuPheProHisLysLeuProArgL 802
|||
QY 2471 TCCCACTGATCCGAGCAGGAGAAACCTTACAGCTCCAGATCTACCCCAACAGAGACAC 2530
|||
Db 802 eupProThrAspProSerArgGluThrLeuProAlaProAspLeuProGlnArgLThrG 822
|||
QY 2531 AGTATTCGCTGCCCCGAGTGGGCGAGCACTA 2562
|||
Db 822 InTyrSerLeuProArgValaGlyArgAlaLeu 832
|||
RESULT 11
ABG61609
ID ABG61609 standard; Protein: 819 AA.
XX
AC ABG61609;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPPP-2 splice variant #8.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPPP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskinestia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
```

```
PN WO200231134-A2.
XX
XX 18-APR-2002.
XX
PE 12-OCT-2001: 2001MO-US31874.
XX
XX 12-OCT-2000: 2000US-240117P.
XX
XX (FERR ) FERRING BV.
XX
PI Qi S, Akinsanya KO, Riviere PI, Junien J;
XX
XX WPI: 2002-444178/47.
XX
DR N-PSDB; ABK83340.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the protein, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
PT
PS Disclosure: Page 98-100; 113pp; English.
XX
XX The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPPP). The dipeptidyl peptidase IV-related proteins (DPPP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychophic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskinestias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPPP proteins.
XX
XX SQ Sequence 819 AA;
XX
XX Alignment Scores:
Pred. No.: 0 Length: 819
Score: 4172.00 Matches: 789
Percent Similarity: 92.61% Conservative: 0
Best Local Similarity: 92.61% Mismatches: 1
Query Match: 87.01% Indels: 62
DB: 23 Gaps: 2
US-09-976-674-4 (1-2617) x ABG61609 (1-819)
QY 11 ATGGCCACACCGGGACCCCAACGGCCGAGCGAGCGACGCGCCGACAGATGACCG 70
|||
Db 30 MetaIatThrThrGlyThrProThrAlaAspArgGlyLysPalaIleAlaIleThrAspAspPro 49
|||
QY 71 GCCCGCCGCTTCCAGGTCGAGAACACACTCGTGGGACGGCTCCGGAGCATCATCCAGGCG 130
|||
Db 50 AlaIleArgPheGlnValaGlnLysHisSerTyrPaspIleLeuArgSerIleIleHisGly 69
|||
QY 131 AGCCGCAAGTACTCGGCGCTCATTTGTATGATCAACAGGCGCCCAACGACTTCCAGTTGTGTCAG 190
|||
Db 70 SerArgLysTyrSerGlyLeuIleValaIleAsnLysAlaIleProHisAspPheGlnPheValaGln 89
|||
QY 251 AGCCGAGAGAACTCCCTCTCTACTGTGATATTCGCAAGAGTCCGGAAGAGGCTCTG 310
|||
Db 110 SerArgGlnLysSerLeuLeuTyrSerGluIleProLysValaIleArgLysGluAlaLeu 129
|||
QY 311 CTGCTCTGTCCTGAGAGCATGCTGATGATTCAGGCGACGCGCCCAACATGAGGCTC 370
|||
Db 130 LeuLeuSerTyrIlePylsGlnMetLeuAspHisPheGlnAlaIleThrProHisGlyVal 149
|||
```


OY	371	TACCTCGGAGGAGGAGCTGCTGAGGAGCAGAAACCCTTGGGGTCTTGGCATCAC	430
Dd	150	TTTserATgglunglungluleuleumarglunglyarsylleucglyvalphedilyllethr	169
OY	431	TCCATGCACCTTCACACGAGAGGGGCTTCCTCTCCAGGCCAGCAACGCTTC	490
Dd	170	Serlyrasphenhissergluserglyleupheupleuhoalasermserrleuhe	189
OY	491	CACGTGCGCAGCGGCGAAGAACGGCTTCATGTTCCCTCATGAACCGGTGGAATC	550
Dd	190	Hiscysargapsglyglylysasn glyPhe metValSerPrometlysproleuendulle	209
OY	551	AAGACCAGAGTCTTAGGGCCCCGGATGGAGCCCAAATGTGCCCTGCGACCTGCCTTC	610
Dd	210	LysThrIngcysSerglyProargMetAspProlysilEcsProAlasproAlaphe	229
OY	611	TTTCCTTCATCATTAACAGCAGCATGTGGTGGCCACAATGAGAAGCGCAGAGAGG	670
Dd	230	PheserPheIleasnAnSerAspLeutrrPvalAlaasnileghnthrlglylueInary	249
OY	671	CGGGTCACCTTCCTCCACCAAGSTTTATCCAATGTCTCGATGAGACCCCAAGTCTCGGT	730
Dd	250	ArgleuthrPhecyshisngln glyLeuserAsnValIleuasPasProlysserlaedly	269
OY	731	GTCGCCACCTTCGTCAATACAGAAAGATTGCACCGCTTCACTGGGTACTGTGTGCCCC	790
Dd	270	ValAlatThrPheValilleglngln lunglunpheasparyPhethrGlytyrTriprypro	289
OY	791	ACACCTTCCTGGGAAGSTTACAGAGGCTTCAGACGCTCGCAATCCTGTATGAGAACTC	850
Dd	290	ThralaserTrpglunglySerglunglyleuLysThrLeuaRyIlleuendryelueInalyal	309
OY	851	GATAGTCCGAGGGGAGGAGCATTCACAGTCCCCTCCGCGCTGGAAGAAAGAGACG	910
Dd	310	AspgluserglvalglueValIlehisValProsePrAlaleuglueInurglysthr	329
OY	911	GACTCGTATCGGTACCCCGAGAGCAGCAGACAATACTCCAGATTGCTTGAATCGACT	970
Dd	330	AspertryrarгыrProargThnglyserLysAsnProLyilEalAleuLysleuAla	349
OY	971	GAGTTCCAGACTGCACACCGGCGAAGTCTGTGCACCCAGAGAAAGAGCTGTGCAG	1030
Dd	350	GluphegntrhrAspsrgnglyn glyLysIleValSerThrgln glylueInalyalgln	369
OY	1031	CCCTTCAGCTGCGCTTCGCCGAGAGGGAGTACATGCCAGGGCGCGGGGACCCGGGAT	1090
Dd	370	ProheserSerleupheProlylsvalglutyrlIleAlaIraPalygllytrPhnarasp	389
OY	1091	GGCAATATACGCGCTGGCCATGTTCTTGACACGGCGCCACAGTGGCTCAGCTGCTCTC	1150
Dd	390	GlylustryalatrPalametPheleuaspargPrognIntrpleuonleuvalleu	409
OY	1151	CTCCCCCGGCGCTTCATTCGCCGAGCAGACAGAGATGAGGAGACCGGCTAGCTTCGC	1210
Dd	410	LeuProAlateleupheIleProseTrhglnasn glylueInalyalaseRala	429
OY	1211	AGACCTGCCCCAGATTCSCAGCGGTATGTGGTATCAGAGAGGTACCAAGCTGCG	1270
Dd	430	ArglatavalProatgasbnValglmProtyrValItylgln glylueInalyalrhAsnalayr	449
OY	1271	ATCAATGTTCATGACATCTTCTATCCCTTCGCCCATCAAGAGGAGAGAGCAGCTGTC	1330
Dd	450	IleasnValHisAspIlePheTytrProPhePrognInserclueglylueAspeluencys	469
OY	1331	TTTTCTCGGCGCAATGAATGCAAGACCGGCTTCGCATTTTACAAAGTCACCGCGTT	1390
Dd	470	Pheleuatrglatasn glyLysThrnglyPhecyshisLeutyryl yValIthralaval	489
OY	1391	TTTAAATCCCAAGGCTCGATTTGAGAGAGCCCTTAGCGCCGGGGAAGATGAATTTAG	1450
Dd	490	LeuLyssereln glylyrsAprrPsegluproPheSerProglylueInaspeInuLheLys	509
OY	1451	TGCCCATTAAGAAGAGATTGCTTGACACGGGTGAATGGAGAGTTTGGCGAGCAC	1510

Db	510	CysProIleLysGluGlnIleAlaLeuThSerGluIuTrGluValLeuAlaArgHis	529
QY	1511	GGSTCGAAGATCTGGTGCATGTAGAGAACAAACTGGTACTTCCAGGGCAACAGAC	1570
Db	530	GlySerLys-----GluThrLysAsp	536
QY	1571	ACGGCGGTGGAGACACACCTCTACGTGTGCATGTATAGAGCGCGCGAGATCGCTACG	1630
Db	537	ThrProLeuGlnHisHisLeuIleuValSerTyrGluAlaAlaGluIleValArg	556
QY	1631	CTCACACAGCGCGGCTTCTCCATTAAGCTGTGCTCCATGAGCGCAACTTGCATGTTCGTC	1690
Db	557	LeuThrTrpProGluYrPheSerHisSerCysSerMetSerGlnAspPheSerPheVal	576
QY	1691	AGCCACGTACACACACGGTACACGCGCGCGCTGGTGGTCCACAGTCTACAGCTAGCGGCCCC	1750
Db	577	SerHisIuYrSerSerValSerThrProProCysValHisValTyrLysLeuSerGluPro	596
QY	1751	GACGACGACCGCGCTCGACACGACGACCGCGCTGTGGGCTAGATGTAGTGGAGCACCAGAC	1810
Db	597	AspAspAspProLeuHisHisLysGlnProArgPheTrpAlaSerMetMetGluAlaIaSer	616
QY	1811	TGCGCGCGGATTAATGTCTCTCCAGAGATCTTCATTTCCACACGCGCTCGGATGTGCGG	1870
Db	617	CysProProAspArgValProProGluIlePheHisPheHisThrArgSerThrAlaArg	636
QY	1871	CTCTACGCGCATGATCTACAGACCGCGCTTGCACGCCAGGACAGACGCCACCGCTC	1930
Db	637	LeuTyrArgMetIleTyrLysProHisAlaLeuGlnProGluLysLysHisProThrVal	656
QY	1931	CTCTTTGTATGTGGAGCGCGCGACGAGTGGACGTGTGAATAACTCCCTGAAGGCATCAAG	1990
Db	657	LeuPheValIuTyrGluGluYrProGlnValGlnLeuValAspAspSerPheLysGluIleLys	676
QY	1991	TACTTGGCGGTCAACACACACTGGCGCTCCCTGGCTACGCGCGTGTGTGATTCAGCGCAGG	2050
Db	677	TyrLeuArgLeuAsnThrLeuAlaSerLeuGluTyrAlaValValIleAspGluArg	696
QY	2051	GGCTCTCTGTACGCGAGGGCTTCGGTTTGAAGGGCGCGCTTAACCAACCAATGGCGCAGGTG	2110
Db	697	GlySerCysGlnArgGluLeuArgPheGluGluAlaLeuLysAsnGluMetGluGlnVal	716
QY	2111	GAGATCGAGGACCGAGTGGAGGGCGTGCAGTTCGTGGCGGACCAATATAGCGTCAATGAC	2170
Db	717	GluIleGlnAspGlnValGluGluLysGlnPheValAlaGluLysTyrGluPheIleAsp	736
QY	2171	CTGAGCGGAGTTGCCATTCATATGGCTGTGCTGCCTACGAGGGCGCTTCTCCGCTCATGGGAGCTA	2230
Db	737	LeuSerArgValAlaIleHisGluTyrPsetTyrGluGluPheLeuSerLeuMetGluLeu	756
QY	2231	ATCCACAAGCGCGGCTGTTCACAAGGTGGCCATTCGCGGGTGGCGCGTCTGATGTG	2290
Db	757	IleHisLysProGlnValPheLys--Ala-----	765
QY	2291	GCGTACGACACAGGGGTCACTGAGCGGCTACATGTGACGTCCCTGAGAACACACGACGCGC	2350
Db	765	-----	765
QY	2351	TATGAGCGGGGTTTCCGTGGCGCTGCACGTGGAGAAAGTCCCAATGAGCCCCAACCGGCTTG	2410
Db	766	-----GlnProLeuAla	769
QY	2411	CTTATCTCCACAGGCTCTCTGAGCAAAAGTGCACATTTTTSCACACAACCTTCCTGCTC	2470
Db	769	IaTyrProProIaArgLeuProGluArgLysArgAlaLeuPheProHisLysLeuProArgL	789
QY	2471	TGCCAATCGATTCGAGCAGGGAACCTTTCACAGTCCCAATATACCCCAAGACAGACAC	2530
Db	789	eupProIAspProSerArgLutThrLeuProIaProAspLeuProGlnArgLutHrg	809
QY	2531	AGTATTCGCTGCGCGCGAGTGGCGCGAGCACTA	2562

DB 809 IntYSerLeuProAlGValGAlYArGAlaLeu 819

RESULT 12
ABG61610
ID ABG61610 standard; Protein; 819 AA.
XX
AC ABG61610;
XX
DT 12-AUG-2002 (first entry)
XX
DE Human DPRP-2 splice variant #9.
XX
KW Human; serine protease; dipeptidyl peptidase IV-related protein; DPRP;
KW DPPIV; infection; human immunodeficiency virus; HIV-1; HIV-2; pain;
KW diabetes; infertility; obesity; anorexia; Parkinson's disease; stroke;
KW heart failure; hypertension; urinary retention; osteoporosis; cancer;
KW ulcer; allergy; cancer; psychotic disorder; neurological disorder;
KW dyskenesia; reproductive disorder; inflammatory disorder;
KW metabolic disorder.
XX
OS Homo sapiens.
XX
PN MO2002J1134-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US31874.
XX
PR 12-OCT-2000; 2000US-240117P.
XX
PA (FERR) FERRING BV.
XX
PI O1 S, Akinsanya KO, Riviere PJ, Junien J;
XX
DR MPI: 2002-444178/47.
XX
PT N-PSDB; ABR83341.
XX
PT New dipeptidyl peptidase IV-related proteins and nucleic acids encoding
PT the proteins, useful for treating e.g. fungal, bacterial, protozoan and
PT viral infections, cancers, allergies, neurological disorders, or pain
XX
PS Disclosure: Page 101-103; 113pp; English.
XX
CC The present invention relates to the isolation of novel human serine
CC proteases referred to as dipeptidyl peptidase IV (DPPIV)-related
CC proteins (DPRP). The dipeptidyl peptidase IV-related proteins (DPRP)
CC and nucleic acids encoding them are useful for treating infections
CC such as fungal, bacterial, protozoan and viral infections, particularly
CC infections caused by human immunodeficiency virus (HIV-1 or HIV-2),
CC pain, diabetes, precocious puberty, infertility, obesity, anorexia,
CC bulimia, Parkinson's disease, acute heart failure, hypertension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC stroke, ulcers, asthma, allergies, cancers, migraine, vomiting,
CC psychotic and neurological disorders (e.g. anxiety, dementia, or
CC schizophrenia), and dyskenesias. These may also be used in discovering
CC therapeutic agents for the treatment of reproductive, inflammatory and
CC metabolic disorders. ABG61591-ABG61612 represent human DPRP proteins.
XX
SQ Sequence 819 AA:

Alignment Scores:

Pred. No.:	0	Length:	819
Score:	4172.00	Matches:	789
Percent Similarity:	92.61%	Conservative:	0
Best Local Similarity:	92.61%	Mismatches:	1
Query Match:	87.01%	Indels:	62
DB:	23	Gaps:	2

US-09-976-674-4 (1-2617) x ABG61610 (1-819)

QY 11 ATGGCCACCAACCGGACCCCAAGCGCGACCGAGGAGCGACCGACCGCCGACAGATGACCGG 70
|||||

DB 30 MetaLaThrThrGlyThrProThrAlaAspArgGlyAspAlaLaLaThrAspAspPro 49

QY 71 GCCGCCGCTTCCAGTGTGCAAGACACTCTGGGACGGCTCCGAGCATCATCAGCGC 130
|||||

DB 50 AlaAlaArgPheGlnValGlnIlyshSerrTrpAspGlyLeuArgSerIleIleHisGly 69
|||||

QY 131 AGCCGCAAGTACTCGGGCGTCATTGTCAACAGAGCGGCCACAGACTTCAGTTTGTCCAG 190
|||||

DB 70 SerArgLysTyrSerGlyLeuIleValAlaShLysAlaProHisAspPheGlnPheValGln 89
|||||

QY 191 AAGACGATAGTCTGGGCCCCCACTCCACCGCTCTACTACCTGGGAAATGCCATAAGC 250
|||||

DB 90 LysThrAspLysSerGlyProHisSerHisArgLeuTyrTrpLeuGlyMetProTyrGly 109
|||||

QY 251 AGCCGAGAGAACTCCCTCTACTGTGAGATTCCCAAGAAAGTCCGGAAGAGGCTCTG 310
|||||

DB 110 SerArgLysLysSerLeuLeuTyrSerGlnIleProLysLysValArgLysGlnAlaLeu 129
|||||

QY 311 CTGCTCTGTCTGCAAGACAGATGCTGATCATTTCCAGGCCACGCCACATGGGCTC 370
|||||

DB 130 LeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 149
|||||

QY 371 TACTCTGGGAGAGAGAGCTGTGAGGAGCGGAAAGCGCTGGGGTCTTCGATACAC 430
|||||

DB 150 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgLysArgLeuGlyValPheGlyTleThr 169
|||||

QY 431 TCCTACGACTTCCACAGCAGAGAGTGGCTCTCTCTCCAGGCGACCAAGCGCTTTC 450
|||||

DB 170 SerTrpAspPheHisSerGlnSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 189
|||||

QY 491 CACTGTGCCAGCGCGGCAAGAACGCTTCATGGTGTCCCTATGAACCGCTGGAATC 550
|||||

DB 190 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGlnIle 209
|||||

QY 551 AAGACCGAGTGTCTGAGGCGCCCGGATGACCCCAAAATCTGCCGCGACCTGCTTC 610
|||||

DB 210 LysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhe 229
|||||

QY 611 TTCTCTCTCATCATTAACAGCAGCTGTGGTGCGCAACATCGAGACGCGAGAGCGG 670
|||||

DB 230 PheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGlnIleGlnGlnIleArg 249
|||||

QY 671 CGGCTGACCTTTGCGCACCAAGGTTTATCCAAATGCTCTGATGACCCCAAGTCTGCGGGT 730
|||||

DB 250 ArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGly 269
|||||

QY 731 GTGGCCACCTTCGTCACAGAGAAAGTGTGACCGCTTCCTGGGTCTGGTGGTGGCCC 790
|||||

DB 270 ValAlaThrPheValIleGlnGlnGlnIlePheAspArgPheThrGlyTyrTrpTrpCysPro 289
|||||

QY 791 ACAGCCTCTGGGAAGGTTGAGAGGGCTCAAGACGCTGCGAATCCTGTATGAGGAATC 850
|||||

DB 290 ThrAlaSerTrpGlnGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlnGlnVal 309
|||||

QY 851 GATGAGTCCAGGTGAGGATTCACGTCCTCTCTGCTAGAGAAAGAGACAG 910
|||||

DB 310 AspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGlnGlnArgLysThr 329
|||||

QY 911 GACTCGTATCGGTATCCCGACAGGACAGGACAGAAAGATCCCAAGTTCGCTGAACCTGGCT 970
|||||

DB 330 AspSerTyrArgTyrProArgTrpGlnGlySerLysAsnProLysIleAlaLeuLysLeuAla 349
|||||

QY 971 GAGTTCAGACTACAGCCAGGCGCAAGATGCTCGACCCAGAGAAAGAGTGTGTCAG 1030
|||||

DB 350 GluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGlnGlyLysGlnValGln 369
|||||

QY 1031 CCCTTACGCTGCTGTTCGCCGAAGGTGAGTACATCGCCAGGCGCGGTGAGACCGGGAT 1090
|||||

DB 370 ProPheSerSerLeuPheProLysValGlnTyrIleAlaArgAlaGlyTrpTrpArgAsp 389
|||||

QY 1091 GCGAATATACGCTGGGCGCATGTTCTGTGAGCGGCGCCAGAGATGGGTCCAGCTGCTCTC 1150
|||||

DB 390 GlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnIleTrpLeuGlnLeuValLeu 409
|||||

QY 1151 CTCCTCCGCGGCTGTTTCATCCGAGCAGAGATGAGAGACGGCTGACCTGCTCC 1210
 |||||||
 Db LeuProProAlaLeuPheIleProSerThrIuAsnGluGlnIuArgLeuAlaSerAla 429
 QY 1211 AGAGCTGTCCCGAGAAATGTCACCCGATATGTGTATGATGAGAGAGTCCACCACTGCTGG 1270
 |||||||
 Db ArgAlaValAlaProAlaGlnValGlnProTyrAlaValTyrGluGlnValThrAsnValTrp 449
 QY 1271 ATCAATGTTCATGACATCTTTATCCCTGCTCCCAATCAGAGGAGAGAGAGAGAGCTGCTGC 1330
 |||||||
 Db IleAsnValHisAspIlePheTyrProPheProGlnSerGluGlnIuAspGlnLeuGlyS 469
 QY 1331 TTTCCGCGCGGCAATGATGACAGACCGGCTTGCCATTTGTCAAAAGTCAACCGCGCTT 1390
 |||||||
 Db PheLeuAlaGlnAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaVal 489
 QY 1391 TTAATAATCCAGGCTAGATTTGAGAGTCCCTTCAGCCCGGAGAGATGAATTTTAA 1450
 |||||||
 Db LeuLysSerGlnGlyTyrAspTrpSerGlnProPheSerProGluGlnAspGluPheLys 509
 QY 1451 TGCCCATTAAGAGAGATTTGCTGTACACAGCGGTGAATGGAGGTTTGGCGAGGAC 1510
 |||||||
 Db CysProIleLysGluGlnIleAlaLeuThrSerGlyIuTrpGluValIleuAlaArgHis 529
 QY 1511 GGCTCCAAAGCTGGGTCAATGAGAGACCAAGCTGCTGCTCCAGGCGACCAAGGAC 1570
 |||||||
 Db GlySerLys-----GlyThrLysAsp 536
 QY 1571 AGCGCGCTGAGACACACCTTACGTGCTAGCTATGAGCGCGCGGCAATGCTGATGCTGACG 1630
 |||||||
 Db ThrProLeuGlnHisHisLeuTyrValValSerTyrGluAlaIleGlnIleValArg 556
 QY 1631 CTCACACAGCGCGCTTCTCCATAGCTGCTCCATGAGCCAGAACTTGACATGTTGCTC 1690
 |||||||
 Db LeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspPheVal 576
 QY 1691 AGCCACTACAGAGGCTGAGACAGCGCGCTGCGGACGCTACAACTGAGCGCGCC 1750
 |||||||
 Db SerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeuSerIlePro 596
 QY 1751 GACGACGACCCCTGACACACAGCCCGCTTGCGGCTTACATGATGAGGACGACGACG 1810
 |||||||
 Db AspAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetMetGluAlaIleAsp 616
 QY 1811 TGCCCGCGGATTTATGTTCTCCAGAGATTTCCATTTCCACACCGCTCGAGTGGCG 1870
 |||||||
 Db CysProProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArg 636
 QY 1871 CTCTACGGGATGATCTTACAGCCCGGCTTGCGGCGGAGGAGAACACCCCGCTC 1930
 |||||||
 Db LeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrVal 656
 QY 1931 CTCTTTGTATGTGAGGCCCGGCTGAGTGTGATTAATCTCTTCAAAAGCATCAAG 1990
 |||||||
 Db LeuPheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyLys 676
 QY 1991 TACTTGCGGCTCAACACACTGGCTCCCTGGGCTACAGCGCTGTGTGATTTGACGGCAGG 2050
 |||||||
 Db TyrLeuAlaGlyLeuAsnThrLeuAlaSerLeuGlyTyrAlaValIleValIleAspLysArg 696
 QY 2051 GGCTCCCTGCTCAGGAGGCTCGGTTGAGAGGCGGCTGAAGAAACCAATGGGCGCAGGTG 2110
 |||||||
 Db GlySerCysGlnArgGlyLeuAlaArgPheGlnGluAlaLeuLysAsnGlnMetCylValIleVal 716
 QY 2111 GAGATCGAGGACAGAGTGGAGGCGCTGACATTCGTGCGCGGAGAAATATGCTTCTCAAC 2170
 |||||||
 Db GluIleGlnAspGlnValGlnGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAsp 736
 QY 2171 CTGACCGGAGTTGCCATCCATGGCTGTCTCTACGGGCGCTTCTCTGCTCATGCGGCTTA 2230
 |||||||
 Db LeuSerAlaGlyAlaIleHisGlyTyrPserTyrGlyGlyPheLeuSerLeuMetCylLeu 756

QY 2231 ATCCAAAGCCCGAGTGTTCAGAGTGGCCATCCGCGGTGCCCGGTCACCGCTGATG 2290
 |||||||
 Db IleHisLysProGlnValPheLys--Ala----- 765
 QY 2291 GCCTACGACACAGGCTACACTGAGCGCTACATGAGACGTCCCTGAGAACACACGACGCG 2350
 |||||||
 Db 765 ----- 765
 QY 2351 TATGAGCGGGTTCGGTGGCCCTGACGTGAGAGAGCTGCCAATGAGCCCAACCGCTTG 2410
 |||||||
 Db 766 -----GlnProLeuA 769
 QY 2411 CTATACCTCCAGGCTTCTGAGCAAAAGCTGCACTTTTCCACAAACTTCTGCTC 2470
 |||||||
 Db IeTyrProProAlaGlnLeuProGlyArgLysArgAlaLeuPheProHisLysLeuProAla 789
 QY 2471 TCCCACTGATTCGAGCAGGAGGAAACCTTACAGGCTCCAGATCTACCCCAAGAGACAC 2530
 |||||||
 Db euProThrAspProSerArgLysIuThrLeuProAlaProAspLeuProGlnArgIuThrG 809
 QY 2531 AGTATTCGCTGCCCGAGTGGGCGAGCACTA 2562
 |||||||
 Db IuTyrSerLeuProAlaGlyAlaArgAlaLeu 819
 Db 809 IuTyrSerLeuProAlaGlyAlaArgAlaLeu 819
 RESULT 13
 AAE23875
 ID AAE23875 standard; Protein: 847 AA.
 XX
 AC AAE23875;
 XX
 DT 23-SEP-2002 (first entry)
 XX
 DE Murine dipeptidyl peptidase 9 (DPP9) protein.
 XX
 KW Murine; dipeptidyl peptidase; DPP; neoplasia; cirrhosis; HIV infection;
 KW human immuno deficiency virus; graft rejection; cytostatic; autoimmunity;
 KW type II diabetes; antidiabetic; antiinflammatory; immunosuppressive;
 KW antiviral; enzyme.
 XX
 OS Mus sp.
 XX
 PN M0200234900-A1.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-AU01388.
 XX
 PR 27-OCT-2000; 2000AU-0001078.
 XX
 PA (UNSY) UNIV SYDNEY.
 XX
 PI Abbott CA, Gorrell MD;
 XX
 DR WPI: 2002-454646/48.
 XX
 DR N-PSDB: AAD38311.
 XX
 PT New dipeptidyl peptidase (DPP) peptides, useful for screening
 PT inhibitors of DPP catalytic activity, which may be employed to treat
 PT e.g. neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 PT rejection and HIV infection -
 XX
 PS Claim 1; Fig 8; 91pp; English.
 XX
 CC The present invention relates to dipeptidyl peptidase (DPP) proteins and
 CC polynucleotides encoding such proteins. The DPP peptides are useful for
 CC screening inhibitors of DPP catalytic activity. The inhibitors are useful
 CC for treating neoplasia, type II diabetes, cirrhosis, autoimmunity, graft
 CC rejection and HIV (human immuno deficiency virus) infection. The present
 CC sequence is an alternative version of murine DPP9 protein.
 CC Note: This sequence is stated to be the same as that shown as
 CC SEQ ID NO: 4 in the sequence listing of the specification. However these
 CC sequences differ.
 CC
 XX

SQ Sequence 847 AA:

Alignment Scores:

Pred. No.:	1.69e-313	Length:	847
------------	-----------	---------	-----

Score:	4129.00	Matches:	764
--------	---------	----------	-----

Percent similarity:	95.79%	Conservative:	32
Best local similarity:	91.94%	Mismatches:	35

```

2000 Local Similarity: 24.24% mismatches: 32
Query Match: 86.11%
Indels: 0

```

```
DB: 23 Gaps: 0
```

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

US-09-976-b/4-4 (1-2617) x AAE23875 (1-847)

[illegible]

Db	310	ArglyrProaagthtrnglySerlySasnpProlysIlaalaIleuylsleuAlaIleuIn	329
Oy	980	ACTGACACAGGGGACAGATGCTCTCGACCCAGAGAGAGACCTGGTGCAGCCTTCACG	1039
Db	330	ThraspnIsngInglyLysIIleValIserserCysGluLysGluIleuValGlnProcheSer	349
Oy	1040	TGCGGTTCGGGAGGAGGTGAGTCAATGCCAGGCGGGGGTGGACCCGGATGGCAAAATAC	1099
Db	330	SerleuPheProlysValGluTyrlIleAlaArgAlaGlyTrpThrArgaspelyLysTytr	369
Oy	1100	GCCTGGGCGCATGTTCTCGAGACCGGGCCAGACACTGGCTCCACTGTCTCTCTCCCGCCG	1159
Db	370	AlatrPlamMetPheIleuAspArgProGlnGlnArgLeuGlnIleuValIleuIleuProPro	389
Oy	1160	GGCCTGTTCAATCCGGACACACAGAAATGAGAGACACCGGCTAGCTCTTCAGAGCTGTC	1219
Db	390	AlaleuPheIleProAlaValGlnSerGluAlaGlnArgIleAlaIleAlaIleAlaArgAlaVal	409
Oy	1220	CCGAGGAATGTCCAGCGGTATGTGGTGTACAGAGAGGTACACAAAGTCTGGATCAATTT	1279
Db	410	ProlysasnValGlnProPheValIleTyrgluGluIleValThrAsnValTrpIleasnVal	429
Oy	1280	CATGACATCTTTATCCCTTCCCTCCCAATCAGAGAGAGACAGACTCTCTTTCTCCGC	1339
Db	430	HisaspIlePheHisProPheProGlnAlaGluGlyGlnIAspPheCysPheIleuArg	449
Oy	1340	GCCAATGAATGCAGACCGGCTCTGCGCATTTGTACAAAGTCAACCGCGTTTAAATCC	1399
Db	450	AlasnGluCysLysTrpThrGlyPheCysHsIleuTyrlArgValThrValGluIleuLysThr	469
Oy	1400	CAGGCGTACGATTTGGATGTGAGCCCTTCAGCGCCGGGAGAAATGAATTTAAAGCCCCATT	1459
Db	470	LysaspTrpIasprTrpThrGlnProIleuSerProThrGluGluIleuPheLysCysProIle	489
Oy	1460	AAGAGAAGATGCTCTGACACACCGGTGAATGGAGGCTTTGGCCAGAGCAGCTCCACAG	1519
Db	490	LysglngluValAlaIleuThrIserGlyGluTrpGluValIleuSerArgHsIleValyseryLys	509
Oy	1520	ATTCGGGCATGTAGAGACACAGACGAGGTACTTGTCCAGGGACACAAAGACAGCGGCTG	1579
Db	510	IletrpValasnGluGlnThrLysIleuValTyrlPheGlnGlyThrLysAspThrProIeu	529
Oy	1580	GAGCACACCTTAAGCTGTGTACACTATAGAGCGCGCGGAGATGTAAGCTCCACACAG	1639
Db	530	GluHsIleuTyrlValIleValIserTyrgIuSerAlaGlyIleValIleValArgIleuThr	549
Oy	1640	CCGGGCTTCGCCATACCTCTCTCCATGAGCGAAGATTGCAATGTGTGTAAGCCATAC	1699
Db	550	LeuGlyPheIserHsIleSerCysIserMetIserGlnIserPheAspMetPheAlaIserHsIstr	569
Oy	1700	AGCAGCGTGAAGACAGCGGCTCGTGACAGCTTACAGCTPAGCGGCGCGAGACAGAC	1759
Db	570	SerIserValIserThrProProCysValHsIleValTyrlLysIleuSerGlyProAspArgsp	589
Oy	1760	CCCCGTGACAAAGACCGCCCTCTCTGGGCTAGCATGATGAGAGACCCAGCTGCCCGCCG	1819
Db	590	ProIeuHsIysGlnProAspArgPheTrpAlaIserMetIserGluAlaIleAsnIysProPro	609
Oy	1820	GATATGTTCTCCACAGATCTCTCAATTTCACACACCGCTGGATGTCCGGCTACAGC	1879
Db	610	AspTrpValIleProProGluIlePheHsIlePheHsIleThrArgAlaAspValGlnIleuTyrcly	629
Oy	1880	ATGATCTACAAAGCCCAAGCCTGTTCAGACCGAGAGAGACACCCACCGCTCTCTTTGTA	1939
Db	630	MetIleuTyrlLysProHsIleThrIleuGlnProGluArgLysHsIleProThrValIleuPheVal	649
Oy	1940	TATGAGAGCCCCAGAGGACGCGGAGGAAATACCTCTTAAAGGATCAAGACTGTGGG	1999
Db	650	TyrlIleuTyrlProGlnValGlnIleuValAsnAspIserPheLysGlyIleuTyrlIleuArg	669
Oy	2000	CTCAACACACTGGGCTCCCTGGGCTACGGCGTGGTGTGTATTCAGAGGAGGGCTCTGT	2059

OY	872	ATTACGTCCTCCCTCCTGGCGCTTGAAGAAGAACAGACGACTGTATGTGGTACCACG	931
Dd	181	IleHisValProSerProIalaLeuGluAlaIlyThrAspSerTyrlarTryProArg	200
OY	932	ACAGGCACCAAGATTCCCAAGATTGCCCTTAACCTGGCTAGTTCCAGATGACGACGGC	991
Dd	201	ThrglSerLysasnProIylsIlleaIaleuLysLeuAlaIupheGlInhrAspserGln	220
OY	992	GCGAAGATCTGTCTGACCCAGGAGAAAGAGACTGTGTGACGCTTCAGCTGGTTCGCG	1051
Dd	221	GlyLysIleValSerThrGlnGluLysGluLeuValGIrproPheSerSerLeuPhePro	240
OY	1052	AAGGGAGGTACATCGCCAGGCGCGGGTGAGACCCGGGATGCGAAATACGCTGGGCGAG	1111
Dd	241	LysValGluTrIleAlaAlaArgAlaGlyTrPrhrArqaspelyLysTyrlalatrPalmet	260
OY	1112	TTCCTGACCGCCCCAGACAGTAGGCTTCACCTGTGCTCTCTCCCGCCGCTGTATCAC	1177
Dd	261	PheIeuasparqprogInglIntprLeuGlnleuValIleueurproIroAlaIeupHele	280
OY	1172	CCGAGCACAGAAATGAGAGACAGCGGCTTAGCTCTGCCAGAGCTGTCCCAGAGATGTC	1231
Dd	281	ProSerThrGluasnGluGluGlnArgLeuAlaSerAlaArgAlaValrProArqsnVal	300
OY	1232	CAGCGGTATGGGTGATCGAGAGAGGTACACACGTCGGATGCAATGTATCATACATCTTC	1291
Dd	301	GlnProTrValValIyTrGluGluValatThrAsnValatrrpleasnValHisSplIephe	320
OY	1292	TATCCCTTCCCCCATCAGAGAGGAGAGAGAGAGAGAGCTGTCTTCTCCGCGCCATACATGC	1351
Dd	321	TyrProPheProgInserGlnGluIyGluAspGluIleucYspheIeuArqAlansngLucs	340
OY	1352	AAGACCGGCTTCCGCTATTGTACAAAGTACCGCGGTTTTAAAATCCAGGGCTACGAT	1411
Dd	341	LysThrGlyPheCySHisLysLeuTyLysValatThrAlaValleuLysSerGlnIyTrysrp	360
OY	1412	TGGAGTGAGCCCTCAGCCCCGGGGAAGATGAATTAAGTGCCCCCATTAAGACAAGATT	1471
Dd	361	TrpserGlnProPheSerProGluGluGly-----GluGln 372	
OY	1472	GCTCTGCACACGGGTGAATGGAGAGTTTGGCAGGACCGGCTCCAAGATCTGGGCTCAT	1531
Dd	373	SerLeuThrAsnAla-----IleThrPalasn 381	
OY	1532	GAGGAGACCAAGCTGCTACTTCCAGAGGACCAAGAGACGCGGCTGGAGACACACCTC	1591
Dd	382	GluGluTrIlyLysLeuValIyPheGlnGlyThrLysAspPhrProIeudInhrShsLeu	401
OY	1592	TACGTGCTACGATTAGAGGGCGCGGAGATGTMCGCTTACACAGCGCGGCTTCTCC	1651
Dd	402	TyrlaValSerTyrguaAlaAlaGlyGluIleValaIArgIeuthrThrProGlyPheSer	421
OY	1652	CATAGCTGCCTCAATGAGCCAGCAACTTGACATTTGGTAGCCACATACAGACGAGCGTGAGC	1711
Dd	422	HisercYsserMetSerGlnAsnPhaeSmePheValaSerHisTyLysSerValSer	441
OY	1712	ACGGCGGCTTGCSTGACSTCTACAAGCTPAGCGGCGCCAGACAGACGCCCCCTGCACAAG	1771
Dd	442	ThrProProcysValHisValIyTrLysLeuSerGlyProAsrpsrpsrProIeunHlys	461
OY	1772	CAGCGCGGCTTGTGGCTAGCATGATGGAGGACAGCCAGCTGCCCGGATATGTCTCT	1831
Dd	462	GlnProArqPrhrArqAlaSerMetMetGluAlaAla----- 473	
OY	1832	CSAAGATCTTCATTTSCACAGCGGCTCGAGTGTGCGGCTACAGGCTGATGTACAAAG	1891
Dd	474	---LysIlePheHisPheHisThrArqSerAsrValaIArgIeuthrIoLyMetIleTyrlus	492
OY	1892	CCCCAGGCTTGGACGCGAGAAAMAGACACCCACGCTCTCTTTGTATATGAGGCCCC	1951
Dd	493	ProHisAlaLeuGlnProGluLysLysHisSPCTChValleuPheValaTyGlyPro 512	

QY	1952	CAGGTGACAGCTGGTGAATTAACCTCTTCAAGACATCAAGTACTGGGTCAACACACTG	2011
Db	513	GlnValGlnLeuValAsnAsnSerPheLysGluIleLysTyrLeuArgLeuAsnThrLeu	532
QY	2012	GCCCTCCGGGGCTACGGCCGGTGGTGGATGGACGGAGGGGGCTCCGTGCAGAGGGCTT	2071
Db	533	AlaSerLeuGluTyrAlaValValValIleAspGluArgGlySerCysGlnArgGlyLeu	552
QY	2072	CGGTTCCGAAGGGGGCTCGAAAAACCAAAATGGGACAGGTGAATGACAGACACAGTGGAG	2131
Db	553	ArgPheGlnGluAlaLeuLeuLysAsnGlnMetGlnGlnValGlnIleGlnAspGlnValGln	572
QY	2132	GAGCTGCAGTTCTGTGGCCGAGAAAGTATGGCTTATCGACCTAGGCGAGGTTGCCATCCAT	2191
Db	573	GlyLeuGlnPheValAlaGlnLysTyrGlnPheIleAspLeuSerArgValAlaIleHis	592
QY	2192	GCGTGGTCTACGGGGGGCTTCCTGCTGGCTATGGGGCTAAATGCACAAACCCAGGTGTC	2251
Db	593	GlyTyrSerTyrGlnGlyPheLeuSerLeuMetGlyLeuIleLysArgGlnValArgHe	612
QY	2252	AAGGGGGCATCGGGGGGTGGCCGGTGCACCGGTGGATGGCTACGACACAGGGTAACT	2311
Db	613	LysValAlaIleAlaGlnArgLysArgProValThrValTyrMetAlaTyrAspThrGlyTyrThr	632
QY	2312	GAGCGCTACATGACACTCCCTCGAAGAACACACACAGCGCTATGAGGGGGTTCCTGGCC	2371
Db	633	GlnArgTyrMetLysArgValProGlnAsnAsnGlnHisGlyTyrGlnAlaGlySerValAla	652
QY	2372	CTGCACCTGGGAGAACTGGCCCAATGACCCACACCGCTTCTTATGCTCCAGGGCTTCG	2431
Db	653	LeuHisValGlnLysLeuProAsnGlnLysProAsnArgLeuLeuIleLeuHisGlyPheLeu	672
QY	2432	GAGCGAAACGTGGCACTTTTTCACACAAACTCTCTGCTCCCACTGATGCGAGACGGG	2491
Db	673	AspGlnAsnValHisPhePheHisThrAsnPheLeuValSerGlnIleLeuValIleGlyAlaGly	692
QY	2492	AAACCTTACACAGCTC-----CAGATCTACCCCAACAGAG	2524
Db	693	LysProTyrGlnLeuGlnValAlaLeuArgProProValSerProGlnIleLysArgAsnGln	712
QY	2525	AGACACAGTATTCGCTGCTCCCGGAGTGGGGGAGACACTATGAAGTCAAGTTCGTGACATT	2584
Db	713	ArgHisSerIleArgCysProGlnSerGlyGlnHisTyrGlnValThrLeuLeuHisPhe	732
QY	2585	CTACAGCAATACCTC	2599
Db	733	LeuGlnGluTyrLeu	737
RESULT 15			
AAM40510			
ID	AAM40510 standard; Protein; 683 AA.		
XX	AAM40510;		
XX	22-OCT-2001 (first entry)		
XX	Human polypeptide SEQ ID NO 5441.		
XX	Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	Leukaemia.		
OS	Homo sapiens.		
XX	WO20015312-A1.		
XX	26-JUL-2001.		
XX	26-DEC-2000; 2000MO-US34263.		

PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI59666.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2: SEQ ID NO 5441: 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA438642-AA442213) with neurotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, Leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 SQ Sequence 683 AA;
 Alignment Scores:
 Pred. No.: 1.62e-267 Length: 683
 Score: 3539.50 Matches: 668
 Percent Similarity: 94.37% Conservative: 3
 Best Local Similarity: 93.95% Mismatches: 37
 Query Match: 73.82% Gaps: 4
 DB: 22
 US-09-976-674-4 (1-2617) x AAM40510 (1-683)
 OY 494 TGTGCGACGAGCGGCAAGACGCGCTTCATGCTGCCCTATGAAACCGGTGAAATCAAG 553
 DB 1 CysAtgAspArgIlylYAsnGlyPheMetValSerPrometIlysProLeuGluIleIys 20
 OY 554 ACCGAGTCTGAGGCGCGCGATGAGACCCCAAAATCTGCCCTGCCGACCTGCTTCTC 613
 DB 21 ThrIlnCysSerGlyProAlaMetAspProIlysIleCysProAlaAspProAlaPhe 40
 OY 614 TCTTCATCATTAACAGGACCTGTGGTGGCAACATCGAGACAGGAGGAGCGGCG 673
 DB 41 SerHeIleAsnAsnSerAspLeuTrpAlaAlaAsnIleGluThrGlyGluIleArgArg 60
 OY 674 CTGACCTTCTGCCACCAAGGTTTATCTCAATGCTCTGATGACCCCAAGTCTGCGGTG 733
 DB 61 LeuThrPheCysHisGlnGlyLeuSerAsnValIleuAspAspProIlySerAlaGlyVal 80
 OY 734 GCGACCTTCGTCATACAGAGAGGTGCAACCGCTTCACATGGGTACTGGTGGCCCCCA 793
 DB 81 AlaThrPheValIleGlnGluIleuPheAspArgPheThrGlyIlyTrpTrpCysProThr 100
 OY 794 GCGTCTGTGGAAGTTTCAGAGGCGCTCAAGACGCTCGAATCCTGTATGAGAACTCAT 853

DB 101 AlaSerTrpGluGlySerGlnGlyLeuIlySerThrIleuArgIleuValIleValAsp 120
 OY 854 GAGTCCGAGGTGAGAGGTATTCACGTCCCTCTGCTGCGCTAGAAAGAAAGACAGCAG 913
 DB 121 GluSerGlyValIleGluValIleHisValProSerProAlaIleuGluIleArgIlyThrAsp 140
 OY 914 TCGTATCGGTACCCGAGGACAGGACGCAAGAAATCCCAAGATGGCTTGAACCTGGCTGAG 973
 DB 141 SerTrpArgIlyTrpAlaGlyThrGlySerIlyAsnProIlyIleAlaIleuIlySerAlaGlu 160
 OY 974 TTCCAGACTGACACGACGAGGCAAGATGCTGTGACCCGAGAGGAGGAGCTGTGACGCC 1033
 DB 161 PheGlnThrAspSerGlnGlyIlySerIleValSerThrGlnGluIlySerGluValGlnPro 180
 OY 1034 TTCACTCTGCTGTTCGCAAGGTGAGTACATCGCCACGCGCGGTGACCCGGGATGCG 1093
 DB 181 PheSerSerLeuPheProIlyValIleGlyTrpIleAlaArgAlaGlyTrpThrArgAspIly 200
 OY 1094 AAATACGCGCTGGCGCATGCTCTGAGCCGCGCGAGAGTGGCTCGACCTGCTCTCTC 1153
 DB 201 LysTrpAlaTrpAlaMetPheIleuAspArgProGlnGlnIleuValIleuValIleu 220
 OY 1154 CCCCCGCGCTGTTCATCCCGAGGACACAGAGATGAGAGCAGCGGCTAGCCTCTGCGAGA 1213
 DB 221 ProProAlaIleuPheIleProSerThrGluAsnGluIleuArgIleuAlaSerAlaArg 240
 OY 1214 GCTGTCCCGAGGAATGTCACCGCTATGTGGTGTACAGAGAGTCCACACGCTGTGATC 1273
 DB 241 AlaValProAlaGAsnValGlnProIlyValIlyTrpGluIleuValThrAsnValTrpIle 260
 OY 1274 AATGTATGACATCTTTCATCCCTTCCCTCCCAATGAGGAGGAGGAGGAGCTGTGCTT 1333
 DB 261 AsnValHisAspIlePheTrpProPheProGlnSerGluGluValAspIleuIlyCysPhe 280
 OY 1334 CTCCGCGGCAATGATGCAAGACCGGCTTCTGCAATTTGTACAAAGTCAACCGCGTCTTA 1393
 DB 281 LeuArgAlaAsnGluCysIlySerThrGlyPheCysHisIleuTrpValThrAlaValIleu 300
 OY 1394 AAATGCCAGGCTACGATGAGAGTACGCTTCACGCGCGGGAAGATTAATTAAGTGC 1453
 DB 301 LysSerGlnGlyTrpAspTrpSerGlnProPheSerProGluGluIly----- 316
 OY 1454 CCCATTAAGAAAGATGTCCTGTGACGCGGCTGAATGGAGGTTTGGCGAGGACGCGC 1513
 DB 317 -----GluGlnSerLeuThrAsnAla----- 323
 OY 1514 TCGAAGATCTGGTTCATGAGAGAGACGCAAGCTGTGACTTCCAGGCGACCAAGACAGC 1573
 DB 324 -----IleTrpValAsnGluGluThrIlySerIleuValIlyTrpIleGlnGlyThrIlyAspTrp 341
 OY 1574 CCGGTGAGACACACCTGTACGTGTACAGTATGAGCGCGCGGCGAGATGCTACGCTTC 1633
 DB 342 ProIleuGlnHisIleuTrpValValSerTrpGluAlaAlaGluIleValArgIleu 361
 OY 1634 ACCAGCGCGGCTCTCCATAGCTGTGCTCCATGAGCCGAGACTTGGATGTGTGTCAGC 1693
 DB 362 ThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSer 381
 OY 1694 CACTACAGACGCTGAGACGCGCGCTGTGCTGCACTGTCAAGCTTACAGGCGCGCGCGAC 1753
 DB 382 HisTrpSerSerValSerThrProProCysValHisValIlyTrpIlySerGlyProAsp 401
 OY 1754 GACGACCCCTGTACACAGACGCGCGCTGTGCGCTACAGATGAGGAGGACGACGCTGC 1813
 DB 402 AspAspProIleuHisIlyGlnProArgPheTrpAlaSerMetSerGluAlaIle----- 419
 OY 1814 CCCCCGATATGCTTCCGAGAGATCTTCATTCGACGCGGTGCGGATGTGGGCGTGC 1873
 DB 420 -----LysIlePheHisPheHisIleThrArgSerAspValArgIleu 432
 OY 1874 TACGCGATGATCTACAGCCCGACGCTGTGACGCGGAGGAGGAGGACCGACGCTCTC 1933

```
Db 433 TyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProthValLeu 452
QY 1934 TTGTATATGAGAGCCCGGAGGTGACCTGTGATATACCTTCAAGGCATCAAGTAC 1993
Db 453 PheValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyr 472
QY 1994 TTGCGGCTCAACACACTGAGGCTCCCTGGGCTACGCCCTGGTGTGATGTAGCGGAGGGGC 2053
Db 473 LeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGly 492
QY 2054 TCCTGTACAGGAGGCTTCGCTCGAAGGGCCCTGAAAAACCAATGGCCAGGTGAG 2113
Db 493 SerCysGlnArgGlyLeuArgPheGlnGlyAlaLeuLysAsnGlnMetGlyGlnValGln 512
QY 2114 ATCGAGGACGAGGTGAGGAGGCTGCATGCTGTGGCCGAGAAGTATGGCTTCATCGACCTG 2173
Db 513 IleGlnAspGlnValIleGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAspLeu 532
QY 2174 AGCCGAGTTGCCATCCATGGCTGGCTCAGGGGGCTTCCTCGCTCATGGGGCTAATC 2233
Db 533 SerArgValAlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerIleuMetGlyLeuIle 552
QY 2234 CACAAGCCCCAGGTGTTCAGAGTGCGCATGGCGGGTCCCGGCTCACCCGTGTGATGGCC 2293
Db 553 HisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValIlePheMetAla 572
QY 2294 TACGACACAGGATACACTGAGCGCTACATGACGTCCCTGAGAACACAGCAGCGCTAT 2353
Db 573 TyrAspThrGlyTyrThrGlnArgTyrMetAspValProGlnAsnAsnGlnHisGlyTyr 592
QY 2354 GAGGCGGGTCCGTGGGCTGACAGTGAGAGAGCTGCCCAATGAGCCCAACCGCTTGCTT 2413
Db 593 GlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeuLeu 612
QY 2414 ATCTCCACAGGCTTCCTGGAGCAAAAGCTCACTTTTCCACACAAACTCTCGTCTCC 2473
Db 613 IleuHisGlyPheLeuAspGlnAsnValHisPhePheHisThrAsnPheLeuValSer 632
QY 2474 CAACTGATCCGAGCAGGAGAACTTACCACTC----- 2506
Db 633 GlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnValAlaLeuProProValSerPro 652
QY 2507 CAGATCTACCCCAACGAGAGACACAGTATTCGCTGCCCGAGTCCGGCGAGCACTATGAA 2566
Db 653 GlnIleTyrProAsnGlnArgHisSerIleArgCysProGlnSerGlyGlnHisTyrGln 672
QY 2567 GTACAGTTGCTGCACCTTCTACAGAAATACCTC 2599
Db 673 ValThrLeuLeuHisPheLeuGlnGlnTyrLeu 683
```

Search completed: December 12, 2002, 11:48:13
Job time : 106 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:04:17 : Search time 18 Seconds
(without alignments)
1410.666 Million cell updates/sec

Title: US-09-976-674-3
Perfect score: 4646
Sequence: 1 MATTGPTADRGDAATDDP.....CPESGHEVTLHLFLOEYL 863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1093.5	23.5	310	US-09-794-236-4	Sequence 4, Appli
2	489.5	10.5	759	PCT-US93-07923-2	Sequence 2, Appli
3	489.5	10.5	766	US-08-230-491A-3	Sequence 3, Appli
4	489.5	10.5	766	US-08-619-280A-3	Sequence 3, Appli
5	489.5	10.5	766	US-08-940-391-3	Sequence 3, Appli
6	489.5	10.5	766	US-09-794-236-1	Sequence 1, Appli
7	482.5	10.4	755	PCT-US93-07923-3	Sequence 3, Appli
8	458	9.9	771	US-09-462-284-2	Sequence 2, Appli
9	426	9.2	760	US-08-230-491A-2	Sequence 2, Appli
10	426	9.2	760	US-08-619-280A-2	Sequence 2, Appli
11	426	9.2	760	US-08-940-391-2	Sequence 2, Appli
12	249	5.4	657	US-09-355-166-1	Sequence 1, Appli
13	240.5	5.2	632	US-09-016-080-1	Sequence 1, Appli
14	207	4.5	593	PCT-US93-07923-11	Sequence 11, Appli
15	163.5	3.5	721	US-09-390-234-20	Sequence 20, Appli
16	163.5	3.5	686	US-09-368-169-8	Sequence 8, Appli
17	160	3.4	622	US-08-664-646A-2	Sequence 2, Appli
18	160	3.4	622	US-09-066-285-2	Sequence 2, Appli
19	160	3.4	622	US-09-261-006-2	Sequence 2, Appli
20	160	3.4	622	US-08-951-088-2	Sequence 2, Appli
21	160	3.4	622	US-09-609-566-2	Sequence 2, Appli
22	160	3.4	622	US-09-609-570-2	Sequence 2, Appli
23	110.5	2.4	255	US-09-355-166-3	Sequence 3, Appli
24	107.5	2.3	346	US-08-602-359A-34	Sequence 34, Appli
25	107.5	2.3	422	US-08-485-938A-34	Sequence 34, Appli
26	106.5	2.3	444	US-08-485-938A-33	Sequence 33, Appli
27	103.5	2.2	614	US-08-446-100-21	Sequence 21, Appli

28	103.5	2.2	614	US-08-446-100-22	Sequence 22, Appli
29	103.5	2.2	614	US-08-446-100-23	Sequence 23, Appli
30	102	2.2	425	US-08-853-659A-38	Sequence 38, Appli
31	101	2.2	809	US-09-186-276B-58	Sequence 58, Appli
32	101	2.2	809	US-08-842-445-58	Sequence 58, Appli
33	101	2.2	809	US-09-186-188B-58	Sequence 58, Appli
34	100	2.2	1382	US-08-737-715-2	Sequence 2, Appli
35	100	2.2	1382	US-09-457-040B-7	Sequence 7, Appli
36	98	2.1	898	US-08-474-379C-12	Sequence 12, Appli
37	98	2.1	898	US-09-146-249A-12	Sequence 12, Appli
38	98	2.1	898	US-08-206-188B-12	Sequence 12, Appli
39	98	2.1	900	US-07-688-352C-12	Sequence 12, Appli
40	97	2.1	657	US-09-370-368-7	Sequence 7, Appli
41	96.5	2.1	870	US-08-785-241-4	Sequence 4, Appli
42	96.5	2.1	870	US-09-374-454-6	Sequence 6, Appli
43	96.5	2.1	963	US-08-537-002A-3	Sequence 3, Appli
44	96.5	2.1	963	US-08-863-010-3	Sequence 3, Appli
45	96.5	2.1	963	US-09-024-429-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-794-236-4
: Sequence 4, Application US/09794236
: Patent No. 6337069
: GENERAL INFORMATION:
: APPLICANT: Grouzmann, Eric
: APPLICANT: Lacroix, Jean-Silvain
: TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
: FILE REFERENCE: 81985/276823
: CURRENT FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 310
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-794-236-4

Query Match      23.5%   Score 1093.5: DB 4: Length 310:
Best Local Similarity 57.3%: Pred. No. 1e+101:
Matches 200: Conservative 42: Mismatches 58: Indels 49: Gaps 1:
```

QY	515	FOGTRDFLEHHLVYVSYEAGEIYRLTPGFSHSCSMSONFDMFVSHYSSVTPCVHV	574
DB	1	FEGTDSPLHHLVYVSYVNGEVTRLDRGSHSCCSQCHDFISXNKNHCVSL	60
QY	575	YKLSGPPDDLPKOPRFNASKMEASCPDPVYPRPIFHFRSDVRLYGMITYKPHALOPG	634
DB	61	YKLSPPDDPCKTKEFATILDSAGPLDYPPIFFESYTGTYLGMITYKPHALOPG	120
QY	635	KKHPTLVLVYGGPOVIVNNFSKIKYLRNLASLGAVVYIDRGSGQRLREGALK	694
DB	121	KKPTVLVLYGGPO-----	134
QY	695	NOMGOVEIEDQEGLOFAEKYGFIDLRSVALHGMSSGFLSLMGLIKPOVFKVAIGA	754
DB	135	---GGIEDDOVEGLQYLAISRYDFIDLRVGHGMSYGYLSLMLMGRSDIFRAVINGA	191
QY	755	PVTVMAYDTGTYERYMDVPPENNQHGYEAGSYALHVEKLPENRLLIHGFLDENVHF	814
DB	192	PVTLMIFDTGTYERYMGHPDQNEQGYIGSYAQAEPSPBNRLLLHGLDENVHFA	251
QY	815	HNFELVSQLIRGKPYQLOIYVNERHSTRCPESGHEVTLHLFLOEYL	863
DB	252	HNSILSLFVRAGKPYQLOIYFOERHSTRVPSGHEVTLHLFLOEYL	300

```
RESULT 2
```

PCT-US93-07923-2
Sequence 2, Application PC/TUS9307923
GENERAL INFORMATION:
APPLICANT: Morimoto, Chikao
APPLICANT: Schlossman, Stuart F.
APPLICANT: Tanaka, Toshiaki
TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
City: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 759
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-2
Query Match 10.5%; Score 489.5; DB 5; Length 759;
Best Local Similarity 22.4%; Pred. No. 4,9e-40;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;
Db 47 LIYKAPHDFOFVOKTDESGPHSHRLYYLGNPYSRENSLLYSEIPKVRKALLLSWK 106
19 VLKNG-----TDDATADSRTYTL-----TDYLNTRYLKLXSLRW- 55
QY 107 QMDHFOATPHGHVYSREELL-----REKRRLGVEGITSYDFH-SSSGLELF 153
56 -----ISDHEVLYYKQENNLVFNAYGNSSVLENSFDESHSINDSISPDGQFIL 108
QY 154 QASNSLFHCRDCKNGKGFVSPMKPLEIKTQCSGPRMPKICPADPAF-FSFINNSDLAWA 212
Db 109 LEVYVYQWHRSTYASVDIDLNKRQLITERIPNNQOWTWSPVGHKLAVVWMDIYVK 168
QY 213 NIEGERRLLTFCHQISLWLDPKSAGVATEVYIOEE-FDRFTGYWCPASWEGSEGLK 271
Db 169 IEDNLPYSRITW--TGKEDIYYN-----GITDWYEEVEFSAYSALMSPNGTF----- 215
QY 272 TLRLIVEVDESEVEYH---VSPALAEKRTDSYRPRGSKPKALKALAEQDTSOG 328
Db 216 ---LATAQENDTEVPLLEYFSYDSDESIOQPYKAVPVPKAGAVPVV--KFEVYVNTDLS 270
QY 329 KIVSTOKELVOPFSSLFPRVEYIARAGWTRDGRYAMAMFLDRPOQWLQVLPPALFIP 388
Db 271 SVYNNATSIQITAPASMLIGD-HYLCDVYTWATQERIS-----LQWL----- 309

QY 389 STENEEOFLASARAVPRNVQPVV-----YEEVTNWVINVHDIPEYPOSGEDELCLFLR 443
Db 310 -----RIQYSVMDICDYDESSGRW----- 330
QY 444 ANECKTGCHLYKXTAYLAKSGYDMSEFSPGDE-----FK 480
Db 331 --NCLVAROHEMSTT-----GWVGRFRSEPHFTLDGNSFYKIIISNEGRHICYFO 381
QY 481 CPIKEEIALTSGEWEVLARHSGKIWNDETKLVYFOGKDTPLBHHLYVSYEAQ---- 536
Db 382 IDKDCFFITGTGTEVIGIEL-----TSD-----YLYISNEYKMGPG 421
QY 537 -EIVRLTTPGFSHSCSMSQNDMEFVSHYSYVTPPCVHVYKL--SGP----- 580
Db 422 RNLKYIOLSDYTKVTYCSCEINPRCOYYSFSGKEAKYQDLRCSGGLPLYTLHSSVND 481
QY 581 -----DDDLPHQOPRFWASMEASCPRPVYPELFFHTBSDVRL-YGMITKPAHQ 632
Db 482 KGLRVLBDNSALDX-----MLQNVQ-----MPSKLDFFILNETKFWYQMLPPLH-FD 528
QY 633 PGKRHPVLYFYGGPOVOLVNNSEFKIKYLRN--TLASLGVAVVV-IDGRGSCQGLR 688
Db 529 KSKYPLLDVYVYAGPCQKADTV-----RLNNATYLASTENIIVASFDRGSGYGDK 582
QY 689 FEGALKNQMGVEIEDQVEGL-QFVAEKYGFIDLSRVAIHGMSYGGELSLMGLIHKPOVF 747
Db 583 IMHAINRRLGTFEEVEDOIEAARQF--SKMGFVDNKRRIAIMGMSYGVYVTSVLGSGGVF 640
QY 748 KVALAGAPVYVMAVYDGYPTFRYMDV--PENQNGYEGSVALVEXKLPNEPNLLIHLG 805
Db 641 KCGIAPAAPSWEYDSDYTERYMGRLPPEEDNDLHYRSTVMSRAENF--KQVEYLLIHG 698
QY 806 FLDENVHFHTNPLVYLSOLIRAGKPYOLOIYPNERHSIRCPESGHEVETLHFLFOE 861
Db 699 TADNVHFQSOAISKRLVYGVDFQALMWTDEBHGIASSTAHQHITTHSHFIKQ 754
RESULT 3
US-08-230-491A-3
Sequence 3, Application US/08230491A
Patent No. 5587299
GENERAL INFORMATION:
APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
APPLICANT: Garlin-Chesa, Pilar; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: FELPE & LYNCH
STREET: 805 THIRD AVENUE
City: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT - ASC II
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5587299man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 766 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-230-491A-3

Query Match 10.5%; Score 489.5; DB 1; Length 766;
 Best Local Similarity 22.4%; Pred. No. 5e-40;
 Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

```

QY 47 LIVNAPHDFOVQKTDSEGRPHSHRLYLGMFGSRENSLSELPKVKRKALLLSMK 106
DB 26 VLNKG-----TDDATADSRKTYTL-----TDYLKNTYRLKLSLRW- 62
QY 107 QMLDHFQATPHHGVSREELL-----REKRRLGVGITSYDFH-SESGFLPF 153
DB 63 -----ISDHEYLTKQENNLIVFNAEYGNSSVFLNSTFDEFGHSINDYSISPDGQFTL 115
QY 154 QASNSLFHCRDGGKNGFVSPMKPLEIKTQCSGPRMDPKICPADPAF-FSFINSDLMWA 212
DB 116 LEVNVYKQMRHSYTASDYIDLNRKRLTEERIPNNQVMTWSPVGHKLAYWNNDIYVK 175
QY 213 NIETGEERLLTFCHOGSLNVLDDPKSAGVATVIOEE-FDRFTGYWCPTASWESGGLK 271
DB 176 IEPNLPYSYRITW--TGKEDIYIN---GITDWYEEVEFSAYSALWMSFNGTF----- 222
QY 272 TLRILYEEDSEVEYIH---VSPALERKTDYRPRTGSKNPKIALKLAEPOTDSOG 328
DB 223 ---LATAQNDVEVPLIEYSFSDSLQPKTVRVYPRAGAVNPV--KFFVYNTDSLS 277
QY 329 KIVSTOEKELVOPFSSLPFKVEYIARAGWTRDGKYAMAMFLDRPOQMLQVLPLPALFTP 388
DB 278 SVTNATSIQITAPASMLIGD-HYLCDVMTWATQERIS-----LQWL----- 316
QY 369 STENEQRLASARAVPRNQPIYV---YEEVTNWVINVHDIETFPPOSSEGEDELCLRL 443
DB 317 -----RRIQNSYVMDICDYDESSGRW----- 337
QY 444 ANECKTGFCHLKVTAVLKSQGYDMSEPPSPGEDE-----FK 480
DB 338 --NCLVAROHIMSTT-----GWVGRPRPSEPHFTLDGNSFYKIISNEGYSRHICYFO 388
QY 481 CPKEEIALTSGEMEVLARHGSKIWNNEETKLVEFGTKDTPLEHLLVYVSEAG---- 536
DB 369 IDKKOCTFTTKGTWEYIGLEAL-----TSD-----YLYISNEYKGMPEG 428
QY 537 -EIVRLTPPGFSHSCSMQNFDMFVSHSVSTPRCVHYKL--SGP----- 580
DB 429 RMLYKIQLSDYTKTCLSCELNPERQYYSVSFSKAKYQYLRCSGPGRLPYTLHSSVND 488
QY 581 -----DDDLHKQPRFNASMMEASCPRDYVPRPEIFHFHRSQVRL-YGMILYKRNALQ 632
DB 489 KGLRVLEDSALDK-----MLQNVQ-----MPSKKLDPIILNEKRYQOMILPRL-ED 535
QY 633 PGKKHPTVLYVGGPOVOLVNNNSFKGIKYLRLN--TLASLGAAYV--IDGRGSCORGJR 688
DB 536 KSKKTPRLILDVYAGCSQADTVF-----RLNMATYLLASTEITIIYASDGRSGYQGRK 589
QY 689 FEGALKNQMGQVEIEDQVBEGL-QFVAEKYGFIDLSHVALHGSYGGFSLMGLIHKPOVE 747
DB 590 IMHAIIRRLGTREVEDEQIBAROF--SKMGFYDNKRRIALIMGSGYGYVSMVLGSSGVF 647
QY 748 KVALIAGARTVMAYDGTGTERIMDY--PENNOHGEAGSVLHVKLNEPRRLILLING 805
DB 648 KCGIAPAVSRWEYDYSYTERIMGLPTPEDNIDHARNSTVMSRAENF--KOVEYLLING 705
QY 806 FLIDENVHFHTNFIQLRACKAPYOLQIYPNRHSIRPESGESEVETLLHPLOE 861
DB 706 TADDNVHFGQSAQISKALVDYGVDFQAMMYTDEDHGIASSTAQHTIYTMASHPIKO 761

```

RESULT 4
 US-08-619-280A-3
 Sequence 3, Application US/08619280A

```

Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Rettig, Wolfgang; Old, Lloyd J.;
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 766 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-3

Query Match 10.5%; Score 489.5; DB 1; Length 766;
Best Local Similarity 22.4%; Pred. No. 5e-40;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

```

```

Db 338 --NCLVANOHIEMSTT-----GWGFRFRSEPHFTLDGNSFKYIISNEGRHICFYQ 388
QY 481 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFGTKDTPLEHHLVVSYEAG---- 536
Db 389 IDKNCCTFTITGWEVIGIEAL-----TSD-----LYYISNEYKGMFG 428
QY 537 -EIVRLTTPGFSHSCSMQNDMFVSHSYSTPCVHYKL--SGP----- 580
Db 429 RNLKIQIOLSDYTKVTCISCELENPERCOYYSFSKKAQYQLRCSGQLPLTYLHSSVND 488
QY 581 -----DDDLHKQPRFWSMMEASCPDYPVPEIFHFHTRSQVRL-YGMITKPHALQ 632
Db 489 KGLRVLEDNSALDK-----MLQNVQ-----MSKKIDFTILNETKFWYQMLPPH-FD 535
QY 633 PGKHPFVLVYGGPOVOLVNSSEFKIKYLRN--TLASLGAVVY-IDRSGSCQGLR 688
Db 536 KSKKYPILLDYAGPCSCAKDTVF-----RLNATYLASTENITVASFDRGSGYGD 589
QY 689 FEGALKNQMGVEIEDOVEGL-QEVAEKYGFIDLSRAIHSYSGFSLMGLIHKPOVF 747
Db 590 IMHAINRRLGTFEVEDQIEAAROF--SKMGFVDMKRIAIMGMSYGGVYTSWVLGSGSGVF 647
QY 748 KVALAGAPVTWMAVDTGTYTERYMDV--PENNOHGEAGSVALHVEKLPNEPNLLIHLG 805
Db 648 KCGIIVAVSVSWEYDYSYTERYMGFLPTPEDNLDHYRNSTVWSRAENF--KQVEYLIIHG 705
QY 806 FLDENVHFHTNLFVSOLIRAGKPYOQIYENRHSIRCPSEGEHYEVLTHFLOE 861
Db 706 TADNVHFOQSAQISKALVDYGVDFQAMWYTDHGHGIASTAHOHITTHMSHFYKQ 761

```

RESULT 5

US-08-940-391-3

Sequence 3, Application US/08940391

Patent No. 5965373

GENERAL INFORMATION:

APPLICANT: Zimmermann, Rainer; Park, John E.;

INVENTOR: Rettig, Wolfgang; Old, Lloyd J.;

TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Felle & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/940,391

CLASSIFICATION: 530

FILING DATE: 01-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/619,280

FILING DATE: 18-MARCH-1996

APPLICATION NUMBER: 08/230,491

FILING DATE: 20-APRIL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5965373man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5330.1

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 766 amino acids

TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-3

Query Match 10.5% Score 489.5; DB 2; Length 766;

Best Local Similarity 22.4%; Pred. No. 5e-40;

Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

```

QY 47 LINVKAPDFOFVQKTDSEGPSHRLYYLGMFYSGRENSLSEIPKKVREALLLSWK 106
Db 26 VLKNG-----TDDATADSRRKTYL-----TDYKNTYRLKLYSLRW- 62
QY 107 QMLDHFQATPHHGYSSREEL-----REKRLGVFGITSTDPH-SSSGFLF 153
Db 63 -----ISDHELYKQENNTLVNAEYGNSSVLENSFDEFGHSINDYSISDGOFIL 115
QY 154 QASNSLPHCRGKNGFVSPMKPLEIKTQCSGRMPKICPADPAF-FSPINSDLVVA 212
Db 116 LEVNTVKMRHSYSTASTDYDLNKRQLTERIPNNTQMTWSPVGHKLAVYWNNDLYVK 175
QY 213 NIETGEERLTFCHQGLSNVLDPKSAGVATFVIOEE-FDRFTGYWMCPTASWEGSEGLK 271
Db 176 IEPNLPYRIW--TGKEDIYN-----GITDYYEEEFSAVSLMSPNGTF----- 222
QY 272 TLRLIYEVEDESEVEVTH--VPSPALEKRTDYSRYPFGSKNPKIALKAEFQDSOG 328
Db 223 --LATAQFNDTEVPLIEYSFYSDESLOYPKTVHVPYKAGAVNPTV--KFFVNTDLS 277
QY 329 KIVSTQEKELVQPSLSLEPKVEYIARAGWTRDGKYYAMAFIDRQOQMLVLLPPLFIP 388
Db 278 SVTNATSIQITAPASMLGD-HYLCDVWATQERIS-----LQWL----- 316
QY 389 STENDEQRLASARAVPRNVQPVV-----YEVTNWNVINDIEFPQSGEDELFLR 443
Db 317 -----RRIQNTSYMDICDYDESSGRN----- 337
QY 444 ANECKTGFCILYKVTAVIKSGDYMSSEFSGEDE-----FK 480
Db 338 --NCLVANOHIEMSTT-----GWGFRFRSEPHFTLDGNSFKYIISNEGRHICFYQ 388
QY 481 CPIKEEIALTSGEWEVLARHSGKIWNNEETKLVYFGTKDTPLEHHLVVSYEAG---- 536
Db 389 IDKNCCTFTITGWEVIGIEAL-----TSD-----LYYISNEYKGMFG 428
QY 537 -EIVRLTTPGFSHSCSMQNDMFVSHSYSTPCVHYKL--SGP----- 580
Db 429 RNLKIQIOLSDYTKVTCISCELENPERCOYYSFSKKAQYQLRCSGQLPLTYLHSSVND 488
QY 581 -----DDDLHKQPRFWSMMEASCPDYPVPEIFHFHTRSQVRL-YGMITKPHALQ 632
Db 489 KGLRVLEDNSALDK-----MLQNVQ-----MSKKIDFTILNETKFWYQMLPPH-FD 535
QY 633 PGKHPFVLVYGGPOVOLVNSSEFKIKYLRN--TLASLGAVVY-IDRSGSCQGLR 688
Db 536 KSKKYPILLDYAGPCSCAKDTVF-----RLNATYLASTENITVASFDRGSGYGD 589
QY 689 FEGALKNQMGVEIEDOVEGL-QEVAEKYGFIDLSRAIHSYSGFSLMGLIHKPOVF 747
Db 590 IMHAINRRLGTFEVEDQIEAAROF--SKMGFVDMKRIAIMGMSYGGVYTSWVLGSGSGVF 647
QY 748 KVALAGAPVTWMAVDTGTYTERYMDV--PENNOHGEAGSVALHVEKLPNEPNLLIHLG 805
Db 648 KCGIIVAVSVSWEYDYSYTERYMGFLPTPEDNLDHYRNSTVWSRAENF--KQVEYLIIHG 705
QY 806 FLDENVHFHTNLFVSOLIRAGKPYOQIYENRHSIRCPSEGEHYEVLTHFLOE 861
Db 706 TADNVHFOQSAQISKALVDYGVDFQAMWYTDHGHGIASTAHOHITTHMSHFYKQ 761

```

RESULT 6

US-09-794-236-1

Sequence 1, Application US/09794236

Patent No. 6337069

```

: GENERAL INFORMATION:
: APPLICANT: Grouxmann, Eric
: APPLICANT: Lacroix, Jean-Silvain
: APPLICANT: Monod, Michel
: TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
: FILE REFERENCE: 81985/276823
: CURRENT APPLICATION NUMBER: us/09/794.236
: CURRENT FILING DATE: 2001-02-28
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO: 1
: LENGTH: 766
: TYPE: PRT
: ORGANISM: Homo sapiens
: us-09-794-236-1

Query Match          10.5%; Score 489.5; DB 4; Length 766;
Best Local Similarity 22.4%; Pred. No. 5e-40;
Matches 201; Conservative 124; Mismatches 330; Indels 241; Gaps 35;

Oy 47 LTVNKAHPDFOVOKTDSGPHSHRLYLGMVYSGRENSLXSELPKVKRKALLLSMK 106
Db 26 VLLNKG-----TDAATADSRKTYTL-----TDYKNTYRLKLSLRW- 62
Oy 107 QMLDHFQATPHHGVYSREBELT-----REKRRLGVGITSYDFH-SESGLELF 153
Db 63 -----ISDHEYLKQENNIILVFNAKEYGNSSVPLENSTFDEFGHSINDYSISPGQFTL 115
Oy 154 QASNSLPHCRDGGKNGFVSPKPLEIKTQCGSPRMDPRICPADPAF-FSFLNNSDLWYA 212
Db 116 LEYNVVKQMRHSYTSYDYLDKRKQLTEERIPNNQWTVWSPGHKLAAYWMNDIYVK 175
Oy 213 NIETGEERLTFCHOGLSVNLDDPKSAGATFVIOE-FDRFTGYWMCPTAEMESGJK 271
Db 176 IEPNLPYRITW--TGKEIDIYN-----GITDWYEEVEVSATLWMSNGTF----- 222
Oy 272 TLRIIYEEVDESEVEYIH---VSPALERKTDYRYPRGSKNPKIALKAEFQDSOG 328
Db 223 ---LVAQFNDEVEPLIEYSFSDSLSQYKTVRVPRYKAGAVNPVY-KFEVYNTDSLS 277
Oy 329 KIVSTQOEKELVOPFSSLPKVEYIARAGWTRDGYAMAFELDRPOQWLOLVLLPALFTP 388
Db 278 SVTNATSIQITAPASMLIGD-HYLCDVWATQERIS-----LCWL----- 316
Oy 369 STENEGOKLASARAVPRNQPYVY-----YEEVTNWINVNDIIFYFPQSEGEDLCFLR 443
Db 317 -----RRIQNTSVMDICDYDESSGRW----- 337
Oy 444 ANECKTGFCHLYKTVAVLKSQGYDWSPPSGEDE-----PK 480
Db 338 ---NCLVAROHIMSTT-----GWVGRPRSEPHFTLDGNSFYKLTISMEGVRHICYQ 388
Oy 481 CPIKEEIALTSGEWEVLARHGSKIWNNEETKLVPFGQTKDPLRHHLLVYVSYEAG- 536
Db 389 IDKDCOTFTKTGWEYIGIEAL-----TSD-----YLYISNEYKMPGG 428
Oy 537 -EIVALLTTFGEFHSMSQSNFDMFVSHYSYSTPRCVNHYKL--SGP----- 580
Db 429 RNLYRIQLSDYTKVCLSCSELNPERCQYVSFSKRAKYUOLRCSGRLPLTYLHSSVND 488
Oy 581 -----DDPLHKQPRFVASMMEAAASCPDYPPEIFHFHRSQVRL-YGMVYKHAHQ 632
Db 489 KCLRVLEDSALDK-----MLOANQ-----MPSKKLDFTILNEKFKYQMLPRH-PD 535
Oy 633 PGKHPYTVLVYGGPQVOLVNNSPKGIKYLRLN--TLASLGAAYVV-IDRGSCQGRGLR 688
Db 536 KSKKPYLLIDYVAGPCQKADTYF-----RLNMATYLASTENIIIVASPDGRSGYQGD 589
Oy 689 FEGALKNQNGUYEIEDQVGL-QFYAEKGFIDLSVALIHGNSYGGFSLMGLIHKPYQF 747
Db 590 IMHAINRRLGTEVEEDQIAARQF--SKMGFVNDKRIALMGWSYGGVYSMLVGSQGVF 647
Oy 748 KVALIGAPVTVMVMAVDTGYTERMDV--DENNOHGYEAGSVALHVEKLPNEPRLLILHG 805

```

```

Db 648 KCGIYAVPSRWEYDYSYTERKMGLPPEEDLDRHNSYVMSRAENF--KQVEYLLHG 705
Oy 806 FLDENVHFHNFNLVSQILRAGKPYQLOIYPRNERSINCPEGHEVETLLHFLQ 861
Db 706 TADNVHFOQSQISKALVDGVDFQAMWYDEDDGIASSTAHOIYTHMSHFIKO 761

RESULT 7
PCT-US93-07923-3
: Sequence 3, Application PC/TUS9307923
: GENERAL INFORMATION:
: APPLICANT: Morimoto, Chikao
: APPLICANT: Schloosman, Stuart F.
: APPLICANT: Tanaka, Toshiaki
: TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM PS/2 Model 502 or 55SX
: OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
: SOFTWARE: WordPerfect (Version 5.0)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07923
: FILING DATE: 19930819
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/934,162
: FILING DATE: 21-AUG-1992
: APPLICATION NUMBER: 07/832,211
: FILING DATE: 06-FEB-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Frazer, Janis K.
: REGISTRATION NUMBER: 34,819
: REFERENCE/DOCKET NUMBER: 00530/055002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 755
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: PCT-US93-07923-3

Query Match          10.4%; Score 482.5; DB 5; Length 755;
Best Local Similarity 22.8%; Pred. No. 2.5e-39;
Matches 194; Conservative 116; Mismatches 321; Indels 219; Gaps 33;

Oy 93 KVKVRKALLLSWKQMLDHFQATPHHGVYSREBELT-----REKRRLGVGITY 140
Db 39 KTYRRLKLSLRW-----ISDHEYLKQENNIILVFNAKEYGNSSVPLENSTFDEFGHS 90
Oy 141 SYDFH-SESGLELFQASNSLPHCRDGGKNGFVSPKPLEIKTQCGSPRMDPRICADPA 199
Db 91 INDYISIPGQFTLELYNVVKQMRHSYTSYDYLDKRKQLTEERIPNNQWTVWSPVG 150
Oy 200 F-FSFLNNSDLWVANIEETGEERLTFCHOGLSVNLDDPKSAGVATFVIOE-FDRFTGYW 257
Db 151 HKIAYWMNDIYKIEPNLPYRITW--TGKEIDIYN-----GITDWYEEVEVSATSLM 204
Oy 258 WCPFTASWESSEGLKTLRIIYEEVDESEVEYIH---VSPALERKTDYRYPRGSKNPK 314
Db 205 WSPNGTF-----LVAQFNDEVEPLIEYSFSDSLSQYKTVRVPRYKAGAVNPY 254

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/230,491A
FILING DATE: 20-APRIL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 558729man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 330
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
FAX: (212) 638-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-230-491A-2

```

```

Query Match          9.2%; Score 426; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 1.3e-33;
Matches 168; Conservative 109; Mismatches 256; Indels 180; Gaps 31;

```

```

QY 195 PADPAFFSTINNSDLVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIOEEDRFT 254
DB 176 PGDPPE-----QITF--NGRENKIFN---GIPWYEEEM-LPT 208
QY 255 GY--WMCPTASWEGSEGLKTLRLIYEVDSEVEVIHVPSPALEE--RKTDSYRYPTGSK 311
DB 209 KYALMWSPPGKF-----LATAEFNDKDIPIVATSYGDEQYPRITINIPYPAGAK 258
QY 312 NPKIALKLAEFQTDGSKIVSTQEKELVQFSSLPKV-----EYIARAGWTRDGKYAMA 366
DB 259 NPVARI-----FIIDTTYPAYVGPQEVVPAMIASDYFSSWLTWTVDKVC-- 305
QY 367 MELDRPOQWLQVLLPPLAFI-----PSTEN--EEORL--ASARAAPRWQPY 410
DB 306 -----LQWLKRVQNVSVSLICDFREDQWTDCKPTQEHIESRTGMAAGFFVSRRVFSY 359
QY 411 VYEEVTNWIVNHIPIFFPQSEGEDELCEFLRANECKTGFCFLKVTAVLKSQGYDWE 470
DB 360 -----DAISYKFIKSDKQYKRIHI----- 380
QY 471 PPSGDEDEKCPIKEETALTSGEMEVLAHSGKIWNNEETKLVYFGQT--KDPLEHILY- 528
DB 381 -----KDTVENAIOITSGKWEAI-----NIFRYTODSLFYSSNEEPEYGRNNIYR 426
QY 529 --VVSYEAGEIIVRLTTPGFSHSCSM-----SONFDMFVSHSVSSTPPCVHYKYL- 577
DB 427 ISIGSYPPSKVCV-----TCHLRKRCQYTTASFSDYAKYVALVCYGPPIPISTLH 477
QY 578 SGPDDEPLH--KQPRWASAMEASCPDPYVPEIHFHTRSDVRLYGMITKYPHALQPGK 635
DB 478 DORTDEIKILEENKLENAKNIOLPKF---EIKKLEVEDITLWYKMLPPO--FDRSK 532
QY 636 KHPYLVFVGGPOVOLVNNNSF--KGIKYLRNLTLASLGYAVVVIDRGSCQORLREGAL 693
DB 533 KRPFLIQVGGCSQSVRSVFANWISYLF---ASKEGVAILVDRGAFQGDKLTYAV 588
QY 694 KNQMGVELEDEQGLQVAAEKYGFIDLSRVAILHGSYGFGLSLMGLIHKPOVFKYALNG 753
DB 589 YKRLGVEVEDQITAVRKFI--NGFIDERRIAITGWSYGYSSLASGTGLFKGCIAY 647
QY 754 APTVYMAVDQVTEHYMVP--ENNQHGVEAGSVALLHKEKLPNEBNRLILHGFIDENY 811
DB 648 APVSWEYVYASVYTERFMGLPTKDDNLEHKNSTVMARAEYFRNV--YLLHGTADNV 705
QY 812 HPHETFLVSOLIRACKPYOLOIYPPNERSIRCPESG---EHYEVTLLHFLQE 861
DB 706 HFGNSQAIKALVNAQVDFQAMWYSDQNGH-----SGLSTNHLTYHTHTPLKO 754

```

RESULT 10
US-08-619-280A-2

```

Sequence 2, Application US/08619280A
Patent No. 5767242
GENERAL INFORMATION:
APPLICANT: Zimmermann, Rainer; Park, John E.;
APPLICANT: Retlig, Wolfgang; Old, Lloyd J.
TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
TITLE OF INVENTION: ALPHA, AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,280A
FILING DATE: 18-MARCH-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5767242man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
FAX: (212) 638-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-619-280A-2

```

```

Query Match          9.2%; Score 426; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 1.3e-33;
Matches 168; Conservative 109; Mismatches 256; Indels 180; Gaps 31;

```

```

QY 195 PADPAFFSTINNSDLVANIETGEERRLTFCHQGLSNVLDDPKSAGVATFVIOEEDRFT 254
DB 176 PGDPPE-----QITF--NGRENKIFN---GIPWYEEEM-LPT 208
QY 255 GY--WMCPTASWEGSEGLKTLRLIYEVDSEVEVIHVPSPALEE--RKTDSYRYPTGSK 311
DB 209 KYALMWSPPGKF-----LATAEFNDKDIPIVATSYGDEQYPRITINIPYPAGAK 258
QY 312 NPKIALKLAEFQTDGSKIVSTQEKELVQFSSLPKV-----EYIARAGWTRDGKYAMA 366
DB 259 NPVARI-----FIIDTTYPAYVGPQEVVPAMIASDYFSSWLTWTVDKVC-- 305
QY 367 MELDRPOQWLQVLLPPLAFI-----PSTEN--EEORL--ASARAAPRWQPY 410
DB 306 -----LQWLKRVQNVSVSLICDFREDQWTDCKPTQEHIESRTGMAAGFFVSRRVFSY 359
QY 411 VYEEVTNWIVNHIPIFFPQSEGEDELCEFLRANECKTGFCFLKVTAVLKSQGYDWE 470
DB 360 -----DAISYKFIKSDKQYKRIHI----- 380
QY 471 PPSGDEDEKCPIKEETALTSGEMEVLAHSGKIWNNEETKLVYFGQT--KDPLEHILY- 528
DB 381 -----KDTVENAIOITSGKWEAI-----NIFRYTODSLFYSSNEEPEYGRNNIYR 426
QY 529 --VVSYEAGEIIVRLTTPGFSHSCSM-----SONFDMFVSHSVSSTPPCVHYKYL- 577
DB 427 ISIGSYPPSKVCV-----TCHLRKRCQYTTASFSDYAKYVALVCYGPPIPISTLH 477

```


OY 578 GCPDDPLH--KOPRFVASMMEASCPDPYVPEIFPHFHTRSVRLYGMIVKPHALOPGK 635
Db 478 DGRDOEIKLIEENKELENKNIQLPKE-----EIKLEVEDITLWKMILPQ-FDRSK 532
OY 636 KHPVLVYGGPOVOLVNNF--KGIKYRLNTLASLGAVVVIDGSCORGLRFEGL 693
Db 533 KYPILIOVYGGPOSVASVAVNMISYL-----ASKEGMIALVDGCTAOGKLLAV 588
OY 694 KNOGVEIEDOVGLOFVAEKYGFIDLSRAVIGHWSYGGFLSMGLTHKPOFVKVAIAG 753
Db 589 YRKLGYEVEDQITAVRKFE-MGFIDEKRIALMGWSYGGVSSALASGTLKCGIAV 647
OY 754 APVTVMAYDGYTERYMDV--ENNCHYAGSVALHVKLPNEPNLLILHGLDENV 811
Db 648 APVSMWEYASVYTERMGTPKTDNLEHKNSTVMARAEYFRVND--YLLIHGTAODNV 705
OY 812 HFFHTNFLVSQIRACKPYOLQIYPERHSIRCPESG---EHYEVTLLHFLQOE 861
Db 706 HFONSAQIAKALVNAQVDFQAMMYSODNHL-----SGLSTNHLTYHMTHTFLKO 754

RESULT 11
US-08-940-391-2
; Sequence 2, Application US/08940391
; Patent No. 5965373
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.;
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
; NUMBER OF INVENTIONS: PROTEIN ALPHA, AND USES THEREOF
; CORRESPONDENCE ADDRESS:
; ADDRESS: 805 Third Avenue
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/940,391
FILING DATE: 01-0CP-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/619,280
FILING DATE: 18-MARCH-1996
APPLICATION NUMBER: 08/230,491
FILING DATE: 20-APRIL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5965373man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5330.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-940-391-2

Query Match 9.2%; Score 426; DB 2; Length 760;
Best Local Similarity 23.6%; Pred. No. 1.3e-33;
Matches 168; Conservative 109; Mismatches 256; Indels 180; Gaps 31;
OY 195 PADPAFFSFINNSDLVANIETGEERLTFCHQGLSNVLDPPKSAQVATFVIOEEDRT 254
Db 176 PDDPP-----QITF--NGRENKIFN---GIPWVYEEM-LPT 208

OY 255 GY--WMCPTASMESEGLKTLRIILEYEVDESEVYIHPSPALBE-RKTDSTYRPRGSK 311
Db 209 KYALMSPNGKF-----LVAEFNDKIDIVIASYVGDOYPTTINIPKAGAK 258
OY 312 NPKIALKIAERQTSOGKIVSTOKELVOPFSSLFPKV-----EYIARAGTIRGKYAMA 366
Db 259 NPVVR-----FIIDTTYPAYVGQEVVPMIASSDYFFSLMTVTERNC-- 305
OY 367 MFLDRPOOMLOLVLLPPALFT-----PSTEN--EQRLL--ASARAVRNVQY 410
Db 306 -----LQMLKRVQNVSVLSICDFREDMQTDWCPKTOEHEIESRTGWAAGGFVSRPFYSY 359
OY 411 VVEEVTNWINVADIFVPPQSGEDICFLRANECKTGCHLYKXTAVLKSQGYMSE 470
Db 360 -----DAISYKLFSPKDKGKTHYI----- 380
OY 471 PESPDEFPKCPIKEIALTNGEWEVLARHGSKIWNEETFLVFOGT-KDTPLEHILY- 528
Db 381 -----KDTVENAIOITSGKWEAL-----NIFRVTQDSLEFSSNPFEEYPCRRNIYR 426
OY 529 --VVSFEAGEIYVLTTPGSHSCSM-----SQNPFMEVSHYSVSTPPCVHYKL- 577
Db 427 ISIGSYPPSKKCV-----TCHLRKRCQYTTASFSDYAKYALVYCGPISITLH 477
OY 578 GCPDDPLH--KOPRFVASMMEASCPDPYVPEIFPHFHTRSVRLYGMIVKPHALOPGK 635
Db 478 DGRDOEIKLIEENKELENKNIQLPKE-----EIKLEVEDITLWKMILPQ-FDRSK 532
OY 636 KHPVLVYGGPOVOLVNNF--KGIKYRLNTLASLGAVVVIDGSCORGLRFEGL 693
Db 533 KYPILIOVYGGPOSVASVAVNMISYL-----ASKEGMIALVDGCTAOGKLLAV 588
OY 694 KNOGVEIEDOVGLOFVAEKYGFIDLSRAVIGHWSYGGFLSMGLTHKPOFVKVAIAG 753
Db 589 YRKLGYEVEDQITAVRKFE-MGFIDEKRIALMGWSYGGVSSALASGTLKCGIAV 647
OY 754 APVTVMAYDGYTERYMDV--ENNCHYAGSVALHVKLPNEPNLLILHGLDENV 811
Db 648 APVSMWEYASVYTERMGTPKTDNLEHKNSTVMARAEYFRVND--YLLIHGTAODNV 705
OY 812 HFFHTNFLVSQIRACKPYOLQIYPERHSIRCPESG---EHYEVTLLHFLQOE 861
Db 706 HFONSAQIAKALVNAQVDFQAMMYSODNHL-----SGLSTNHLTYHMTHTFLKO 754

RESULT 12
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1

Query Match 5.4%; Score 249; DB 4; Length 657;
Best Local Similarity 20.6%; Pred. No. 7.3e-16;
Matches 152; Conservative 112; Mismatches 246; Indels 228; Gaps 34;
OY 179 EIKTQSGP-----RMDPKICPADPAFFSFINN-----SDLVANITETGEERLTFCH 226
Db 50 EIKTGSVWTHGEKRSSTDRMSP-DGRLATFSDREGAAQQLIYINSTGGEGRKRLTIDP 108
OY 227 OGISNVLDPPKSAQVATFVIOEEDRFTGYWMCPTASMESEGLKTLRI-LYEEVDESEV 285

```

Db 109 YGVSKPL-----WSP-----DGSIIYITISLGESESIDR-- 138
QY 286 EVIHPSPALERKTDSTRYPTGSKNPKIALKLAEPQIDSG-----KIYST 333
Db 139 -----KKTQDSDYE-----PVEQGLSYKRDGGLTRGAYADLVVSVKSG 179
QY 334 QEKELVOPFSSLPFVEYIARAGWTRDGKYAMAMFLDRPQOMQLVLLPPLALFISTPENE 393
Db 180 EMKELTIS-----HKADH-GDPAPFSPDGK--WLVF-----SAULT 210
QY 394 EORLASAPAVRNVOPIVYVEEVTNWMINVDIIFYPPQSGEDELCEFLRANECKTGFC 453
Db 211 ETDAS-----KPHDVY-----IMLSGLDK----- 232
QY 454 LYKTAVALKSGCYDMSEFSP-----GEDEKCPKEKRIALTSGEW----- 494
Db 233 -----QVTPHKGSGFSGSSFPDGRYLLALGNEKEYK-----NATLSKAMLYDIBQRLT 281
QY 495 ---EVLARH-----GSK---IWNVEETKLYVFOGTRKDPPLREHNLVVSYEAG 536
Db 282 CLTEMLDVHLADALIGDLIGAGBQRPIW-TKDSGCFYVIGDQG--STGIYIISIEGLV 338
QY 537 EIVRLTFPFSHSCSMQNFDMFVSHSYSTPFCVHVYKLSGDDPLHKQPRFMA5MM 596
Db 339 YPIRLEK-YINSFSLSPDEQHFIA5YTKPRDPS--ELYSI-----PLQGEKOLGTGAN 389
QY 597 EAASCPPIVPELIFHEHTRSDYRLXGMITKPHALQPKKHPTVLFGVGGPOVLVNSF 656
Db 390 DKFVREHTISIPETIOYATEDGVNWMGLMRPOMEGETTYPLILINHGPIIMMGHYTF 449
QY 657 KGIKYRLNTLASLGYAVAVVIDGRSCORGLFEGALKNMQOVEIEPOVEILOFAEKY 716
Db 450 H-----EFOVLAAKYIAVYVYINPRSGHSGOEFVNAVRGDYDGKRYDVMQAVDAIKRD 504
QY 717 GTIDLSVAIHGWSYGGFLS--LMGLHKPOYKVALAGAPYVWMAV---DTGY--TE 768
Db 505 PHIDKRLGVTGSGYSGFMTNWIYQOTNR---FKAAYTORISNMISFSGVSDIGYFED 561
QY 769 RYM--DVEENNQHGVEASV--ALHVEKLPNPNRLLIHLGGLDENVHEFHNFVLSQL 824
Db 562 WOLEHDMEDVEDEKLMRSPKLYAANVE-----TPLLILHGERDRCPLEQAOLFLALK 615
QY 825 RAGKPYOQIYFNERHSI 842
Db 616 KMGKETKLVPRPNASHNL 633

```

RESULT 13
US-09-016-080-1
; Sequence 1, Application US/09016080
; Patent No. 6133012

; GENERAL INFORMATION:
; APPLICANT: Ishikawa, Kazuhiko

; APPLICANT: Matsui, Ikuro

; APPLICANT: Ishida, Hiroyasu

; APPLICANT: Kosugi, Yoshiyugu

; APPLICANT: Higuchi, Katsuhiko

; TITLE OF INVENTION: THERMOSTABLE ACTYL PEPTIDE HYDROLASE AND GENE ENCODING

; FILE REFERENCE: 07898/022001

; CURRENT APPLICATION NUMBER: US/09/016,080

; EARLIER FILING DATE: 1998-01-30

; EARLIER FILING DATE: 1997-01-31

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 632

; TYPE: PRT

; ORGANISM: Pyrococcus horikoshii

; US-09-016-080-1

Query Match 5.28; Score 240.5; DB 4; Length 632;
Best Local Similarity 20.8%; Pred. No. 4.9e-15;
Matches 158; Conservative 105; Mismatches 272; Indels 225; Gaps 36;

```

QY 132 KRLGVFGITTSDFHSE---SGFLFQASNSLFFHCHDGSKNFPWSPMKPLEIKQCSOPR 188
Db 23 KGAIVQVTEISLKDDDFSKLYLY-----DGKR-----VKPFTSGMKN5MR 65
QY 189 MDPKICPADPAFPFSFIN-----NSDLVANIENGEEERRLTFCHQGSNV--LDDPKSA 239
Db 66 FSP-----NGKLIATFSKRDKGKESLEYLPTDGGKRLAKRKYGIKILRFTEDGKSI 120
QY 240 GVAT-----FVIOEFDRFTGYWMCPRASWEGSLKTLRLIYEVESEVEY 287
Db 121 AVVPIPIDEKKGNDVHLIRLPEFNGVW-----IYGR--RNMYVL 161
QY 288 IHVPSPALERKTDSTRYPTGSKNPKIALKLAEPQIDSGCK IYST--QEKELVOPFSS 345
Db 162 VDV-----ESGKKRRLTPKN---LNVDOIRFHN--GRLYFTAOEDRERKPLISDL 206
QY 346 F-----PKVEYIARAGWTRDGKYAMAMFLDRPQOMQLVLLP--PALFIPSTENEBORLAS 399
Db 207 YLENRRKVRKLT-----DKMRLDPLDQSGFVLAKNLTLE----- 243
QY 400 ARAVPRNVOPIVYVEEVTNWMINVDIIFYPPQSGEDELCEFLRANECKTGFCILKYTA 459
Db 244 -RGITPNH-----IYHDP-----KTG--ELKRLTK 267
QY 460 VLKSGCYDMSEFSGEDELCEFKPIKEETALTSGEMEVLARHSGKTIWNEETKLYVFG-- 517
Db 268 DLDRAVNV-----SLNSDVRSQRAELVYKEGMI 296
QY 518 ---TKDTPLEHNLVYVSEAGEIYRLTPGSHSCSMQNDMFSVSHSYSTPFCVHV 574
Db 297 YVATDGP--RAVLFVRVNDL--GKIERVTL--GDSRVSEFSDIGDYLAFTAOAVTTELYI 351
QY 575 YKLSGDDPLHKQPRF--WASMEASCPPIVPELIFHEHTRSDYRLXGMITKPHALQ 632
Db 352 YR-----DGEKKKYVDFPKMVIKGYTLK-----PEHKVYKASGVDELDAVMKPVNFR 399
QY 633 PEKKHPTVLFGVGGPOVLVNSFSGIKYL--RLNTLASLGYAVVVIDGRSCORGLRREG 691
Db 400 KGKYPALLEIHGPKTAY-----GYAFMEHFHVLTSKGFVYI5NPRSGSDGYDEEF--A 452
QY 692 ALKNQMGVEIETDOYEGLOFAVNAKGYGTIDLSVAIHGWSYGGFLS--LMGLHKPOYK 749
Db 453 DIRGHYGERDYDLMVDEALRRDFIDGERLGTGSGYSGFMTNWIYQOTNR---FKA 509
QY 750 ATAGAPYVWMAV---DTGY---TERYMDVPEENNQHGVEAGSVLAHVEKLPNEPNRLLI 802
Db 510 AVTQRSISNMISFEGTIDIGYFAPDQIGKDPMSNLEGYWEKS---PLKYPANVEFTPLLI 566
QY 803 LHGFLDENVHEFHNFVLSQLIRAGKPYOQIYFNERHSI 842
Db 567 IHSTEDYRCMLPEALQFLISKYLIGKRVLAIFPENHDL 606

```

RESULT 14
PCT-US93-07923-11
; Sequence 11, Application PC/TUS9307923

; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao

; APPLICANT: Schlossman, Stuart F.

; APPLICANT: Tanaka, Toshiaki

; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

; NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:
STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02110-2804

```
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07923
FILING DATE: 19930819
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,162
FILING DATE: 21-AUG-1992
APPLICATION NUMBER: 07/832,211
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00530/055002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 593
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US93-07923-11
```

```
Query Match 4.58; Score 207; DB 5; Length 593;
Best Local Similarity 18.9%; Pred. No. 1.1e-11;
Matches 136; Conservative 93; Mismatches 255; Indels 234; Gaps 31;
```

```
QY 47 LIVKAPHDFQVQKTESGPHSHRIYLGMPYSGRENSLLYSEIPKVKREALLLSMK 106
DB 26 VLLKKG-----TDDATADSKRTYTL-----TYLKNTYKLKLSLRW- 62
QY 107 QMLDFQATPHHGVSRSEELL-----REKRLGVFGITSDFH-SESGLEPLF 153
DB 63 -----ISDHEIYKQENNLIVNAEYGNSSVLENTSPFEGHSINDYISPDGQFIL 115
QY 154 QASLSLPHCRDCKNGENGVSPMKLEIKTQCSGPRMDPKICPADPAF-FSFINNSDLWA 212
DB 116 LEYNYVQWHRSHSYASYDIYDLNKRROLITEERIPNNTQWTVMSPVGHKLAVYMNNDIYVK 175
QY 213 NIEGERRRLTFCHOGISLVNDDPKSAGVATFVIOE-FDFFTGYVMCPTSMGSEBKL 271
DB 176 IEPNLPSTRIWT--TGKEDIIYN---GIDWVYEEVFSAVSLMWSPNCTF----- 222
QY 272 TLRLIYEVEDESEVEVH---VPSPALEERKTDGYRPRTSKNPKTALKLAEFQDSOG 328
DB 223 ---LAVQFNDETEVPLEIYSFYSDPSLQYKTVAPYPRKAGAVNPTV--KFFVNTDLS 277
QY 329 KIYSTOKELVOPSSLEPKVEYIARAGWTRDGKYAMAMFLDRPOQLOLVLLPALFIP 388
DB 278 SVTWTATSIQITAPASMLIGD-HYLCDTWATQERIS-----LQWL----- 316
QY 389 STENEBRILASARAVPRNVQPYV---YEEVNTVMINVDIYFPFQSGEDELCLRL 443
DB 317 -----RRIONYSVMDICDYDESSGRW----- 337
QY 444 ANECKTGFCHLYKVTAVLKSQGYDMSPEFSGEDE-----FK 480
DB 338 --NCLVAROHIEKMT-----GWGFRFRSEPHFTLDGSPFKIITSNENGYHICFYQ 388
QY 481 CPIKEEIALTSGEWEVLARHGSKITWNEETKLYVFOGKTDPLEHLLVYVSEAGA--- 536
DB 389 IDKDCFTFITKGTWEVIGIEAL-----TSD-----YLYYISNEKGMHGG 428
QY 537 -EIVRLTTPGFSHSCSNQFDMFVSHYSYSTPCVHYVYL--SGP----- 580
DB 429 RNLYKIQLSDTYKVTCLSCELNPERCOYYSVSFSKAKRYQLRCSGGLPLYLTHSSVND 488
```

```
QY 581 -----DDDLPHKOPRFWASMEASCPDYVPELTFHFTSRDVL--YGMIXKPHALQ 632
DB 489 KGLRVLEDNSALDK-----MLQNVQ-----MPSKIDFIILMETKRWYMLLPH-FD 535
QY 633 PGKHPVLYEVYGGPOVOLVNNSEKIKYLRN--TLASIGYAVV-IDGRGSCQRG 686
DB 536 KSKRYPLLDYVAGPSCQKADTVF-----RLNMATYLASTENIIVASFDRGSGYQG 587
```

RESULT 15

```
US-09-390-234-20
Sequence 20, Application US/09390234
Patent No. 6365390
GENERAL INFORMATION:
APPLICANT: Blum, David L.
APPLICANT: Kateeva, Irina
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Phenolic Acid Esterases, Coding Sequences and Methods
FILE REFERENCE: 67-98
CURRENT APPLICATION NUMBER: US/09/390,234
EARLIER FILING DATE: 1999-09-03
EARLIER APPLICATION NUMBER: US 60/099,136
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 721
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-09-390-234-20
```

```
Query Match 3.58; Score 163.5; DB 4; Length 721;
Best Local Similarity 21.5%; Pred. No. 3.7e-07;
Matches 98; Conservative 56; Mismatches 193; Indels 107; Gaps 20;
```

```
QY 467 DWSEP-FSPGDE-----FKCPIKEEIALTSGEWEVLARHGSKIW 505
DB 289 DSSSPVFPNDKLAIPQMDTEYSDRALLYVSLGSKTIIIPVAGDWD--ASPDPV- 344
QY 506 VNEETKLYVFOGTYDTPLEHLLVYVSEAGEIVRLTTPGFSHSCSNQFDMFVSHYS 565
DB 345 -----KWTLP-DGKTLIYVSEDLGRTLRFLSPANAKDDYPRKNF---TDGGS 386
QY 566 VSTPPCVHYVYKLSGRDDPLHKOPRFWASMEASCPDYV-----P 607
DB 387 VSA-----YFLL--PDSSLVTGSALMTNMVYTAAPKEGVYIKKIASANEIDPELKLGP 439
QY 608 PEIHFHTRSD-VRLYGMITYKPHALQPGKHPVLYVGVGPQVOLVNNSEFKIKYLRN- 665
DB 440 SDISEFYFGNFTIHAMVYIPENFDSKRYPLFFIHGQ-----GMMADGNS-TRNRP 494
QY 666 -TLASIGYAVV-IDGRGSCQRLEFGALKNQMGVEIEDOVESGLQVAEKYYGIDLSRV 724
DB 495 KAMADOGVVVAPRPTGSTFGQALTTAIONMGCAPYDDLVKCEYVHENIADVTDHG 554
QY 725 AIHGMSTYGGFL-----SLMGLHKPQYFK--VAIAGAPV---YMAAYDGYTERINDV 773
DB 555 VAAAGSTGEGFINMIGQSPGAKRKALVSHDGTVAADKYSTBELWP-MOREFPGTWD 613
QY 774 PENNQHGEAGSVVALHEKLPNEPNR-----LLILHGFIDENVHFTNFTLVSLIRA 826
DB 614 RDN-----YRRND-----PSAPERILQFTPYRLVJHSDQDYLRPAVEGLSLFNVLQER 661
QY 827 GKPYOLOIYPERNHSIRCPESGEHYEVTLLHFOEY 862
DB 662 GVPSPRFLNFPDENHVVVNPENSLVHOOALGWINKY 697
```

Search completed: December 12, 2002, 12:07:17
Job time : 24 secs.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: December 12, 2002, 11:41:37 ; Search time 72 Seconds
(without alignments)
14978.481 Million cell updates/sec

Title: US-09-976-674-4
Perfect score: 4795
Sequence: 1 caagctacatgagccacca.....tctgagcgagcgagatccg 2617

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p_model -DEV=xlh
-O=cg92.1/USPTO.spool/US09976674/runtac_04122002_162359_5987/app_query.fasta.1.2759
-DB=SPRMBL_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09976674.ecgn_1.1_94_etunat_04122002_162359_5987 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPRMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvivirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query Match Length DB ID Description
No. Score
1 4646 96.9 863 4 Q8WXD8 Q8WXD8 homo sapien

2	2887.5	60.2	883	11	Q9DAG6	Q9d4g6 mus musculi
3	2870	59.9	882	4	Q9HBM5	Q9hbm5 homo sapien
4	2572	53.6	508	4	Q75273	Q75273 homo sapien
5	2383	49.7	439	4	Q9BVR3	Q9bvr3 homo sapien
6	2137.5	44.6	632	4	Q96JX1	Q96jx1 homo sapien
8	1859.5	38.8	432	4	Q75868	Q75868 homo sapien
9	1675	34.9	312	4	Q96NWT8	Q96nwt8 homo sapien
10	1631.5	34.0	1042	5	Q9VC20	Q9vc20 drosophila
11	1631.5	34.0	1102	5	Q9VCI9	Q9vci9 drosophila
12	1494.5	31.2	465	4	Q9HBM3	Q9hbm3 homo sapien
13	1376	28.7	469	4	Q9NKF4	Q9nxf4 homo sapien
14	1220.5	25.5	360	4	Q9HBM4	Q9hbm4 homo sapien
15	1093.5	22.8	310	4	Q9HBM4	Q9hbm4 homo sapien
16	943.5	19.7	927	5	Q965F3	Q965f3 caenorhabdi
17	934.5	19.5	931	5	Q4987	Q4987 caenorhabdi
18	834.5	17.4	746	10	Q9RNF6	Q9rnf6 arabidopsis
19	805.5	16.8	738	16	Q9AE0	Q9ae0 caulobacter
20	776	16.2	741	2	P95782	P95782 xanthomonas
21	585.5	12.2	711	2	Q47900	Q47900 flavobacter
22	543	11.3	723	2	Q66223	Q66223 porphyromon
23	534	11.1	723	2	Q31048	Q31048 porphyromon
24	511	10.7	901	3	Q96V77	Q96v77 aspergillus
25	508.5	10.6	730	2	Q93JY4	Q93jy4 prevotella
26	505.5	10.5	793	3	Q14073	Q14073 schizosacch
27	486	10.1	765	6	Q8WNG8	Q8wng8 bos taurus
28	479.5	10.0	799	5	Q9N217	Q9n217 felis silve
29	471.5	9.8	748	13	P70092	P70092 xenopus lae
30	469.5	9.8	765	3	Q14425	Q14425 aspergillus
31	465	9.7	707	16	Q9F348	Q9f348 streptomyces
32	458	9.6	771	3	Q42812	Q42812 aspergillus
33	457	9.5	804	11	Q92218	Q92218 mus musculi
34	453	9.4	755	13	Q91651	Q91651 xenopus lae
35	449.5	9.4	803	11	Q9QVY8	Q9qv8 ratius sp.
36	447.5	9.3	988	5	Q9VMB4	Q9vmb4 drosophila
37	444	9.3	711	4	Q9P236	Q9p236 homo sapien
38	438.5	9.1	761	11	Q8R492	Q8r492 ratius norv
39	430.5	9.0	237	2	Q87543	Q87543 capnocytoph
40	409.5	8.5	802	5	Q9VUH1	Q9vuh1 drosophila
41	405.5	8.5	795	16	Q9PHC9	Q9phc9 xylella las
42	370	7.7	1349	4	Q8WMO4	Q8wmo4 homo sapien
43	368	7.5	745	5	Q9VWM2	Q9vwm2 drosophila
44	360.5	7.5	772	2	Q9AN37	Q9an37 bradyrhizob
45	359	7.5	789	2	Q93M73	Q93m73 xanthomonas

ALIGNMENTS

RESULT 1
ID Q8WXD8 PRELIMINARY; PRT; 863 AA.
AC Q8WXD8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Dipeptidyl peptidase 9.
GN DPP9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Olsen C., Wagtmann N.;
RT "Identification and characterization of a novel member of the
RT dipeptidyl peptidase IV-related family";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF452102: AAL47179.1: -
DR InterPro: IPR002469: DPPV_N-term.
DR InterPro: IPR001375: Peptidase_S9.
DR InterPro: IPR002410: Pro_aminopeptase.
DR InterPro: IPR000379: Ser_estrs_site.
DR Pfam: PF00930: DPPV_N-term; 1.


```

QY 1991 TACTTGGCGCTCAACACACATGCGCTCCCTGGGCTACGCGCTGTGTGATTGACGGCAG 2050
    |||||||
Db 661 TTTTLeuATgLeuAenThLeuAlaSerLeuGlyTrrAlaValValIleAspLyrArg 680
QY 2051 GCGTCTCTCAGCGAGGCTTCGGTTCGAGAGGGCCCTGAAAACCAATGGGCCAGGTG 2110
    |||||||
Db 681 GlycerGlnArgGlyLeuAArgPheGlnGlyAlaLeuLysAsnGlnMetCylValVal 700
QY 2111 GAGATCGAGGACGACGAGGAGGCGCTGAGTTCGTCGGCCGGAACATATGCTTCAATCAG 2170
    |||||||
Db 701 GlnTleGlnAspGlnValGlnGlyLeuGlnPheValAlaGlnLysTyrGlyPheIleAsp 720
QY 2171 CTGAGCCGAGTTCGATCCATGCTGCTCTACGCGGCTTCCTCTCCCTCATGGGCTA 2230
    |||||||
Db 721 LeuSerATgValAlaIleIleGlyTrrPserTyrGlyPheLeuSerLeuMetCylLys 740
QY 2231 ATCCACAAAGCCCGAGCTTCAAGGTGCGCATCGCGGGTGCCTCCGCTCTGATG 2290
    |||||||
Db 741 ILeHISLysProGlnValAlaPheLysValAlaIleAlaGlyAlaProValThrValTrpMet 760
QY 2291 GCGTACGACACAGGCTACACTGAGCGGCTACATGACGCTCCCTGAGAAACACAGCAGCGC 2350
    |||||||
Db 761 AlATyrAspTrpGlyTyrThrGlnArgTyrMetAspValProGlnAsnGlnHisGly 780
QY 2351 TATGAGCGGGTCCGCTGCGCTGACGCTGAGAGAGTGCCTCAATGAGCCCAACCGCTTG 2410
    |||||||
Db 781 TyrGlnAlaGlySerValAlaLeuHisValGlnLysLeuProAsnGlnProAsnArgLeu 800
QY 2411 CTATATCTCCACAGGCTTCCTGAGCAAAAGCTGACATTTTTCACACAAATCTCTCTC 2470
    |||||||
Db 801 LeuIleLeuHisGlyPheLeuAspGlnLysValHisPheHisThrAsnPheLeuVal 820
QY 2471 TCCCAACGATACCCGAGCGGAGAAACCTTACACGCTCCGATATACCCCAACAGAGACAC 2530
    |||||||
Db 821 SerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnArgHis 840
QY 2531 AGTATTCCTGCTCCGAGCTGCGGCGAGCAGCATATGAAGTCAGCTTCGATCTTCTACAG 2590
    |||||||
Db 841 SerIleATgCysProGlnSerGlyLysIleHisTyrGlnValThrLeuLeuHisPheLeuGln 860
QY 2591 GAATACCTC 2599
    |||||||
Db 861 GlnTyrLeu 863

RESULT 2
ID Q9D4G6 PRELIMINARY; PRT; 883 AA.
Q9D4G6
AC Q9D4G6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE 4932434F09RIK.
GN Mus musculus (Mouse).
OS Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090.
RX NCBI_TaxID=10090.
RA SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=TESTIS;
RA MEDLINE=21085660; PubMed=11217851.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Botfeill D., Boujona N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,

```

```

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nodone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016546; BAB30295.1; -
DR MEROPS; S09.018; -
DR MGD; MGI:1921638; 4932434F09RIK.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estr_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 883 AA; 101297 MW; EA958D6BD33EC71F CRC64;

Alignment Scores:
Pred. No.: 3,83e-184 Length: 883
Score: 2887.50 Matches: 520
Percent Similarity: 76.72% Conservatve: 139
Best Local Similarity: 60.54% Mismatches: 197
Query Match: 60.22% Indels: 3
DB: 11 Gaps: 3

US-09-976-674-4 (1-2617) x Q9D4G6 (1-883)
QY 26 ACCCCACGCGCCGAGCGAGCGAGCGAGCGCGCGAGATGACCGCGCGCGC---TTC 82
    |||
Db 17 ThrAlaGlnCysGlnGlnGlyLysGlnGlyLysGlnSerGlnAspArgProLysLeuGlnProPhe 36
QY 83 CAGGTGCGAGAACACTGCTGGGAGCGGCTCCGAGACATCATCCAGCGAGCGCGAGATAC 142
    |||
Db 37 TyrValGlnAlaGlySerTyrSerGlnLeuLysLysLeuValAspThrArgLysTyr 56
QY 143 TCGGCGCTCATGTGTCAACAGCGCGCCGACGACTTCGATTTGTGCAAGAGCGGATGAG 202
    |||
Db 57 HisGlyTyrMetMetAlaLysAlaProHisAspPheMetPheValLysArgThrAspPro 76
QY 203 TCTGGCGCCACCTCCACCGCGCTCTACTACTCGGATGCCATATGCGAGCGGAGAGAAC 262
    |||
Db 77 AspArgProHisSerAspArgValTyrTyrLeuAlaMetSerGlyGlnAsnArgGln 96
QY 263 TCCCTCTCTACTGTGAGATTCCCAAGAGTCCGGAAGAGCTCTGCTCTCTCTCC 322
    |||
Db 97 ThrLeuPheTyrSerGlnIleProLysThrIleAsnArgAlaValLeuMetLeuSer 116
QY 323 TGGAGCAGATGCTGATCATTTCCAGCGCGAGCCGCCACCATGGGGGTACTCTCGGGAG 382
    |||
Db 117 TrpLysProLeuLeuAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArgGln 136
QY 383 GAGGAGCTGCTGAGGAGAGCGGAAAGCGCTGGGGTCTTGGCATCACTCTACGACTTC 442
    |||
Db 137 GlnGlnLeuLeuArgGlnLysArgIleGlyThrValGlyIleAlaIleArgLysPyr 156
QY 443 CACACGAGAGTGGCTCTCTCTCTCCAGCGCGAGCAAGCGCTCTCCACTGTGGCGAC 502
    |||
Db 157 HisProGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisIleLysAsp 176
QY 503 GCGCGCAAGAACGCGCTTCATGGTGTCCCTATGAACCGCTCGGAATCAAGACCCAGTSC 562
    |||
Db 177 GlyGlyProHisGlyPheThrGlnGlnProLeuAlaGProAsnLeuValGlnThrSerGly 196
QY 563 TCAGGCGCCGAGATGAGCCCAAAATCTGCCTCCGCGAGCCCTCTCTCTCTTCATC 622
    |||
Db 197 ProAsnIleArgMetAspProLysLeuGlyProAlaAspProAspTrrPleAlaPheIle 216
QY 623 AATACAGGAGACTGTGGTGGCCAAACATCGAGACAGGAGAGAGCGGGCGCTACCTTC 682
    |||
Db 217 HisSerAsnAspIleTrrPleSerAsnLeuValThrArgGlnGlnArgArgIleThrTyr 236
QY 683 TCCCAACGAGGTTTATCCAAATGCTCTGATGACCCCAAGTCTGGCGGTGGCCACCTTC 742

```

```

Db      237 ValHisnsgluLeuAlaAsnmeLgluLnsprOArGserAlaIValAlaIAlaThrPhe 256
OY      743 GTCAATACAGAAAGATTGCGACGCTTACAGTACGTGGTGGTCCCGACAGCCGCTCGG 802
Db      257 ValLeuIngluLnsprArGlySerLysTrpTrpCysProGlnAlaGluArg 276
OY      803 GAAGGTTCAAGAGGCGCTCAAGACGCTGGAATCCTGTATGAGAAGTGCATGAGTCCGAG 862
Db      277 ThrProserGlyLys--LysIleLeuAlaGlyIleLeuTrgIuGluAAsnArGluSerGlu 295
OY      863 GTGAGGTCAATCACTGCTCCCTGCTGCGCTGATACAGAAGGAAGACGAGCTATCCG 922
Db      296 ValGluIleIleHisValThrSerProMetLeuGluThrArgAlaAspSerPheArg 315
OY      923 TACCCAGACGAGCGACAGAAATCCCAAGATGCTTGAACGTGGGCGATGCTTCAGACT 982
Db      316 TyrProLysThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleValAla 335
OY      983 GACAGCCAGGCAAGATCGTCTGACCCAGAGAGAGAGAGAGTGTGACGCCCTTCAGCTCG 1042
Db      336 AspAlaAlaGlyIleIleAspValIleAspLysGluLeuValGlnProPheGluIle 355
OY      1043 CTGTCCCGAAGGTGAGTATCATGCCAGGCGCGGTGGAGCCGGATGCGAAATACGCC 1102
Db      356 LeuPheGluGlyValGlyTrpIleAlaArgAlaGlyTrpThrProGluGlyLysHisAla 375
OY      1103 TGGGCGCATGCTCGACGCGGCGCCCGACAGTGGCTCCAGGCTGCTCCGCTCCCGCGGCC 1162
Db      376 TrpSerIleLeuLysArGlySerGlnThrHisLeuGlnIleValLeuIleSerProGlu 395
OY      1163 CTGTTCATCCCGACAGACAGAAATGAGAGAGAGAGAGCGCTAGCGCTTCGACAGCTGTCCC 1222
Db      396 LeuPheIleProValGluLnsprAlaMetAspArgGlnArgLeuIleGluSerValPro 415
OY      1223 AGGAATGTCCAGCGCTATGAGTGTACAGAGAGAGTCAACAAGCTGTGATCATGTTCCAT 1282
Db      416 AspSerValThrProLeuIleIleTrpGluIuThrThrAspIleTrpIleAsnIleHis 435
OY      1283 GACATCTTATCCCTCCCGCCCAATCAAGAGAGAGAGAGAGAGTGTGCTTCCGCGGCC 1342
Db      436 AspIlePheHisValRheProGlnThrHis--GluAspGluIleGluPheIlePheAla 454
OY      1343 AATGAATGACAAGACGCGCTGTCCATTTGTACAAGTCAACGCGGTTTAAATCCAG 1402
Db      455 SerGluLysTrpTrpGlyPheArgHisLeuTrpLysIleThrSerIleLeuGlyGluSer 474
OY      1403 GGCATGATGGAGTGAGCCCTTCAGCGCGGGAAGAATTAATTAATGCGCCATTAAG 1462
Db      475 LysTrpLysArgSerSerGlyLysLeuProAlaProSerAspPheLysCysProIleLys 494
OY      1463 GAAGAGATGCTGTGACAGCGGTGAATGGAGGTTTGGCGAGAGACGAGCTCCAAAGT 1522
Db      495 GluGluIleThrIleThrSerGlyLysIuTrpIleValLeuGluArgHisGlySerAsnIle 514
OY      1523 TGGGTCAATGAGACAAAGAGTGGTACTTCCAGGCGACCAAGAAGACGCGCTGGAG 1582
Db      515 TrpValAspGluAlaArgLysLeuValTrpPheGluGlyThrLysAspSerProLeuGlu 534
OY      1583 CACCACCTTACAGTGTGACAGTATGAGCGCGCGGAGATCGTACGCTCACACAGCCC 1642
Db      535 HisHisLeuTrpValThrSerTrpAlaAsnProGlyLysValAlaArgLeuThrAspArg 554
OY      1643 GGCCTTCCCATAGCTGCTGCATGAGCGACAAATCTTCACATGTTCCGAGCATACAGC 1702
Db      555 GlyTrpSerHisSerCysLysSerArgHisLysAspPheIleSerLysTrpSer 574
OY      1703 AGCGTGAAGACGCGCGCTGCTGACAGTCAACAGCTGAAGCGCGCCGAGACGAGCCCC 1762
Db      575 AsnGlnLysAsnProHisCysValSerLeuTrpLysLeuSerSerProGluAspArgPro 594
OY      1763 CTGCACAAGAGCCCGCTTGTGGGCTAAGCATGATGAGGAGCGACGCTGCCCGCGGAT 1822
:::|||||
:::|||||

```

```

Db      595 ValHisLysThrLysGluPheTrpAlaThrIleLeuAspSerAlaGlyProLeuProAsp 614
OY      1823 TATGTCCTCCAGAGATCTTCATTTCCACAGCGGCTCGGATGTGCGGCTGTACGGCATG 1882
Db      615 TyrThrProProGluIlePheSerPheGluSerThrThrGlyPheThrLeuTrpGlyMet 634
OY      1883 AATCAAGGCGCCAGCGCTGCTCAGCCAGGAGGAAGACGCCAGCCGCTTGTGTATAT 1942
Db      635 LeuTrpLysProHisAspLeuGlnProGlyLysLysTrpProThrValLeuPheIleTrp 654
OY      1943 GAGAGCGCCCGAGGTCAGCTGTGATTAATCTTCACAAAGCATCAAGTACTGCGCTC 2002
Db      655 GlyLysProGlnValGlnLeuValAsnAsnArgPheLysGlyValLysTrpPheArgLeu 674
OY      2003 AACACATGCGCTCCCTGCGGTACGCGGTGGTGTGATGTCGAGAGAGCGGCTGTGACG 2062
Db      675 AsnThrLeuAlaSerLeuGlyTrpValValAlaValIleAspAsnArgGlySerCysHis 694
OY      2063 CGAGGCTTGGGTTGGAAGGCGCCTGAATAAACCAATGGCGCAGTGTGAGATGAGAGAC 2122
Db      695 ArgGlyLeuLysPheGluGlyAlaPheLysTrpLysMetGlyGlnIleGluIleAspArg 714
OY      2123 CAGGTGAGGCGCTGCACTTCGTGCGCGAGAAATGATGAGTTCATGCACTGAGCGCGATT 2182
Db      715 GlnValGluGlyLeuGlnTrpLeuAlaSerGlnTrpAspPheIleAspLeuAspArgVal 734
OY      2183 GCGATCCATGAGCTGTGCTCAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2242
Db      735 GlyIleHisGlyTrpSerTrpGlyGlyLysLeuSerLeuMetAlaLeuMetGlnArgSer 754
OY      2243 CAGGTGTTCAGGTGGCCATCGCGGCTGCGCGGCTGCAACGCTGTGATGCGCTTACAGACA 2302
Db      755 AspIlePheArgValAlaIleAlaGlyAlaProValThrLeuThrIlePheTrpAspThr 774
OY      2303 GCGTACACTAGCGCTACATGAGAGCTCCCTGAGAAACAACAGACGCGCTATGAGCGGCT 2362
Db      775 GlyTrpThrGluArgTrpGlyMetGlnHisProAspGlnAsnGluGlnGlyTrpLeuGly 794
OY      2363 TCCGTGCGCGCTGCAAGTGTGAGAGAGTGCSCAAATGAGGCGCAACGCTGTGATGCTCCAC 2422
Db      795 SerAlaIleMetGlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuHis 814
OY      2423 GCGTCTCGACAAAGAGTGCACATTTTCCACAAATCTTCGTCGCTCCCAATGATC 2482
Db      815 GlyPheLeuAspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuVal 834
OY      2483 CGACAGGAGAAACCTTCACAGTCAAGTCAATGAGGCGCAACGCTGTGATGCTCCGCTGC 2542
Db      835 ArgAlaGlyLysProTrpAspLeuGlnIleTrpProGlnGluArgHisSerIleArgVal 854
OY      2543 CCCGAGTGGGCGAGCACTATGAAGTCAAGTGTGCTGCACTTCTACAGAAATACCTC 2599
Db      855 ProGluSerGlyGlnHisTrpGluLeuHisValLeuLeuHisLysTrpLeuGlnAlaSerLeu 873

```

RESULT 3

```

O9HBM5
ID O9HBM5 PRELIMINARY; PRT; 882 AA.
AC O9HBM5:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-JUN-2002 (TrEMBLrel. 16, Last sequence update)
DE Dipeptidyl peptidase 8.
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Plimates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA Gottell M.D.;
RA "Cloning, expression and chromosomal localization of a novel human

```


OY	2183	GCACATTCATGGCTATGCTCTACGGGGGCTTCCTGCTCATGGGCTATCCAAAGCC	2242
Db	301	AlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetClyLeuIleHisLysPro	320
OY	2243	CAGGTGTTCAGAGTGCGGCATCCGCGGGTCCCGGCTACCTGTGGATGAGCTACGACACA	2302
Db	321	GlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAspPhe	340
OY	2303	GGGATCACTGAGGCTCATGAGACGTCCCTGGAGAACACACACAGGGCTATGAGGGGGGT	2362
Db	341	GlyTyrThrGlnAlaGlyTyrMetAspValProGlnAsnAsnGlnHisGlyTyrGlnAlaGly	360
OY	2363	TCCGTGGCCCTGCACATCGTAGAAGACTCCCAATGAGCCCAACCGCTGTATTCTCCAC	2422
Db	361	SerValAlaLeuHisValGlyLysLeuProAsnGlnProAsnArgLeuIleLeuHis	380
OY	2423	GGCTCTCTGGACCAAAACGTCGCACTTTTTCACACAAATCTCTGTCTCCCACTATGC	2482
Db	381	GlyPheLeuLysProGlnAsnValHisPhePheHisThrAsnPheLeuValSerIleLeuIle	400
OY	2483	CGACACAGGAAACCTTTCACAGCTCCAGATCTACCCCAACGAGACACAGATTCGGCTGC	2542
Db	401	ArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGlnLysHisSerIleArgLys	420
OY	2543	CCGCAATCGGGGACGACACTATGAAAGTCAGCTTGCTGCACATTTCTACAGAGATACCTC	2599
Db	421	ProLysSerGlyGlnHisTyrGlyValAlaThrLeuLeuHisPheLeuGlnGlyTyrLeu	439
RESULT 6			
096JX1			
ID	096JX1	PRELIMINARY;	PRF: 632 AA.
AC	096JX1		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	CDNA FLJ14920 fis, clone PLACE1007416, weakly similar to dipeptidyl		
DE	peptidase IV (EC 3.4.14.5).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PLACENTA;		
RA	Isoeqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shirafori A., Sudo H.,		
RA	Magatsuma M., Hosoki T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,		
RA	Nakamiya K., Iwayanagi T.;		
RT	"NEDO human cDNA sequencing project."		
SL	Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AK027826; BAB55395.1; -.		
DR	MORPIS: S09.018; -.		
DR	SEQUENCE 632 AA: 72639 MW; 9BDF598B06985DA4 CRC64;		

Alignment Scores:	
pred. NO.:	3 59e-134
Score:	2197.50
Percent Similarity:	72.30%
Best Local Similarity:	58.52%
Query Match:	44.58%
DB:	4
	Gaps: 3
	length: 632
	Matches: 395
	Conservative: 93
	Mismatches: 144
	Indels: 53
	Gaps: 3

US-09-976-674-4 (1-2617) x Q96JX1 (1-632)

QY 575 ATGACCCCAAAATCTGCCCTGCCGACCCCTCTTCTTCCTTCATCAATACAGGCAC 633
|||||:::|||||::: |||::: |||::: |||
Db 1 MetaspProlySleucYspProIaAspProIaAspPrrIleAaPheIIeHisSerAsnasp 20
QY 635 CTGTGGGTGGCCCAATCTGACACACAGCGAGGAGCGCGCGCTGACCTTCTGCCACCAAGT 694

Db	21	IlErfpIleSerAsnIlleValtThArgGluGluArgGluThrTyraIHisAsnGlu	40
QY	655	TTATTCATGTCCTGGATGACCCCAAGTCGCGGGGTGGCCACCTTGCTCATACAGGA	754
Db	41	LeuAlaAsnMetGluGluAspAlaArgSerAlaGlyValAlaIleThrPheValLeuGlnGlu	60
QY	755	GAGTTCGACCGCTCACCTGGGGTACTGGTGTGGCCCCACAGCGCTTCGGGAAGGTTCCAG	814
Db	61	GluPheAspArgTyrseryGlyTYTTPTPCysProLysAlaGluThrThrProserGly	80
QY	815	GGCCTCAGACGCTGCCAATCTGTATGAGGAAGTCGATGAGTCCGAGGTGAGGTCATT	874
Db	81	Gly--LysIleLeuArgIleLeuTyrgLugLAsnAspLysSerGluValGluIleIle	99
QY	875	CACGTCCCTCTCTCGGCTTAGAAGAAAGAAACAGGACTCGATTCGGTACCCAGACGA	934
Db	100	HisValThrSerProMetLeuGluThrArgAlaAlaAspSerPheThrTyrseryIle	119
QY	935	GGCAGCAAGATCCCAAGATTGGCTTGGAACCTGGCAGAGTCCAGACTACAGACCGAGC	994
Db	120	GlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGluGly	139
QY	995	AAGATCGTCTGCACCCAGAGAAAGAGACTGTGTCAGACCCCTCAGCTCGCTTTCCGAG	1055
Db	140	ArgIleIleAspValIleAspLysGluLeuIleGlnProPheGluIleLeuPheGluGly	159
QY	1055	GTGGATGATCATCGCCAGCGCGGGGTGACCCGGGATGCGAAATAGCCCTGGGCCATGTT	1114
Db	160	ValGluTyrlleAlaIleArgAlaGlyTPThrProGluGlyLysTyraIleThrserIleLeu	179
QY	1115	CTGAGCGGCGCCAGCAGTGGCTCCAGCTGTCCTCTCTCCCGGCTGTTCATCCCG	1174
Db	180	LeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGluLeuPheIlePro	199
QY	1175	AGCAGACAGTGAAGGAGGAGCGGCTAGCCTCGCCAGAGCTGTCCCGAAGTCCAG	1234
Db	200	ValGluAspAspValIleMetGluArgGluIleGluIleGluSerValProAspserValThr	219
QY	1235	CCGATGCTGTATCGAGGAGGTACCAACGTCGTGATCAATGTTTCATCTTCTAT	1294
Db	220	ProLeuIleIleTyrgLugIuThrThrAspIleTrpIleAsnIleHisAspIlePheHis	239
QY	1295	CCCTCCGCCCATAGAGGAGGAGAGACACTGCTTCTCCGGCCCAATGAATGACAG	1355
Db	240	ValPheProGlnSerHis--GluGluGluIleGluPheIlePheIleAspIleLysIle	258
QY	1355	ACCGGCTTCTCCATTTGTATCAAAAGTCACCGCGCTTTAAATCCCAAGGCTACAGATTGG	1414
Db	259	ThrGlyPheArgHisIleLeuTyrsIleIleThrSerIleLeuLysGluSerTyrsIleAsp	278
QY	1415	AGTGAAGCCCTTCAGCCCCGGGGAAGATGAATTTAAAGTCCCCCATTAAGAGAGATGCT	1474
Db	279	SerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGluIleIleAla	298
QY	1475	CTGACACGCGGTGAATGGAGGATTGGTGGCAGGACGAGGCTCCAAAGTTCGGGTCAATGG	1534
Db	299	IleIleThrserGlyIuTrpGluValLeuGlyAlaHisGlySerAsnIleGluValAspGlu	318
QY	1535	GAGACCAAGCTGTGTACTTCCAGGGGACCAAGAGACGCGCGTGGAGCACCACCTTAC	1594
Db	319	ValArgArgLeuValTyrsPheGluGlyThrLysAspSerProLeuGlnHisIleLeuTy	338
QY	1595	GTGGTCACGATATAGAGGGCGCGCGCGAGATGTATGCGCTTACCAACGCGCGCTTCCAT	1654
Db	339	ValValSerTyraIleAsnProGlyGluValThrArgLeuThrAspArgSerTyrseryHis	358
QY	1655	AGCTGCTCATATGACCCGAACTTCGACATTTGCTGACGCTACAGCAGCCTGAGCAGC	1714
Db	359	SerTyrsCysIleSerGlnHisIleCysAspPhePheIleSerLysTyrsSerAsnIleLysAsn	378
QY	1715	CCGCGCTTCGTCACGTCATCAAGCTGAGCGGCGCCCGACGAGCAGCCTTCGCAAGCAG	1774

```

Db 379 ProHiscysValserLeuTyrlsLeuSerProGluAspProThrCysLysThr 398
QY 1775 CCCCCTCTGGCTAGATGATGAGCACCACCTCCGCCGATTAATGTTCTCCA 1834
Db 399 LysGluPheThrPalatHrIleLeuAspSerAlaGlyProLeuProAspTyrThrProPro 418
QY 1835 GAGATCTTCCATTTCCACACGCGCTGATGGCGCTGAGGCTACAGCTACACAGCC 1894
Db 419 GluIlePheSerPheGluSerThrThrIlePheThrLeuTyrlGlyMetLeuTyrlsPro 438
QY 1895 CACGCGCTTGCAGCAGGAGAGACACCCACCGCTCTTGTATATGAGGCCGCCAG 1954
Db 439 HisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyProGln 458
QY 1955 GTGCAGCTGGTGAATACCTCTTCAAGCATCAAGTCTGGCGCTCAACACACTGGCC 2014
Db 459 ValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrIleuAla 478
QY 2015 TCCCTGGGCTACGCGCGTGTGTGATGATGACGAGGCGCTCGTACAGGAGGCTTGG 2074
Db 479 SerLeuGlyTyrValValValValValIleAspAsnArgLysSerCysHisArgGlyLeuLys 498
QY 2075 TTGCAAGGGGCGCTGAAAAACCAATGGCCAGGTGAGATCGAGACCAAGTGGAGGCG 2134
Db 499 PheGluGlyAlaPheLysTyrLysMet----- 507
QY 2135 CTGCAGTTCTGGCGGAGAAAGATATGGCTTCATCGACTGAGCCGAGTTGCCATCCATGCG 2194
Db 507 ----- 507
QY 2195 TGGTCTACAGGGGGCTCTCTCGCTCATGGGGGCTAATCCACAGCCCGAGGTTTCAAG 2254
Db 507 ----- 507
QY 2255 GTGGCCATCGGGGCTGCCCGGTCACCGTCTGGATGGCTACGACAGAGGTACACTGAG 2314
Db 508 ValAlaIleAlaGlyAlaProValThrLeuThrPheThrLysArgPheGlyThrGln 527
QY 2315 CGCTACATGACGCTCCCTGAGAACACACGACGACGCTATGAGGGCGGTTCCGTGGCCCTG 2374
Db 528 ArgTyrMetGlyHisProAspGlnAsnGlnGlyTyrTyrLeuGlySerValAlaMet 547
QY 2375 CACGTGAGAGAGCTGCCAATGAGCCCAACCGCTTATTCACAGGGCTTCTCTGGAC 2434
Db 548 GlnAlaGlnLysPheProSerGluProAsnArgLeuLeuLeuHisGlyPheLeuAsp 567
QY 2435 GAAAAAGTGAAGCTTTTCCACACAAACTTCTCGTCCCAACTGATCCGAGCAGGAGAA 2494
Db 568 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaGlnGlyLys 587
QY 2495 CCTTACACAGCTCCAGATGATACCCCAAGACAGACAGACATGCTGCCCGCAGTGGGCG 2554
Db 588 ProTyrAspLeuGlnIleTyrProGlnLysArgHisSerIleArgValProGluSerGly 607
QY 2555 GAGCAGTATGAAGTCAAGCTTGGTCACTTCTACAGAAATACCTC 2599
Db 608 GlnHisTyrGluLeuHisLeuLeuHisTyrLeuGlnGluAsnLeu 622

RESULT 7
ID 075868 PRELIMINARY: PRT: 432 AA.
AC 075868:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE R33083.1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

```

```

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangaran L., Ertler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Atlix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carraro A.V.;
RT "Sequence analysis of a 2.5 kb region in 19p13.3.";
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC005783; AAC62840.1;
FT NON_TER 432
SQ SEQUENCE 432 AA; 48595 MW; 64E2B85BE0523A7E CRC64;

Alignment Scores:
Pred. No.: 1,07e-115 Length: 432
Score: 1859.50 Matches: 354
Percent Similarity: 97.52% Conservative: 0
Best Local Similarity: 97.52% Mismatches: 0
Query Match: 38.78% Indels: 9
DB: 4 Gaps: 2

US-09-976-674-4 (1-2617) x 075868 (1-432)
QY 11 ATGGCCACACCGGGAGCCCAACGCGCAGGAGCGAGCGCCGACAGATGACCG 70
Db 72 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaThrAspAspPro 91
QY 71 GCCGCCGCTTCCAGGTGACAGAACGACATCGGGAGCGGCTCCGAGCATCATCCAGCG 130
Db 92 AlaAlaArgPheGlnValGlnIleHisSerThrAspArgGlyLeuArgSerIleIleHisGly 111
QY 131 AGCCGCAAGTACTGGGCTCTATTGTCAACAAGCGCCCGACGACTTCCAGTTTGTGAG 190
Db 112 SerArgLysTyrSerGlyLeuIleValAlaAsnLysAlaProHisAspPheGlnPheValGln 131
QY 191 AAGAGGATGAGTGGGCGCCGACATCCGACGCGCTCTACTCGTGGGAATGCGATATGCG 250
Db 132 LysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 151
QY 251 AGCCGAGAACTCCCTCTCTACTCTGATGATTCGCAAGAGTCCGGAAGAGGCTCTG 310
Db 152 SerArgGluAsnSerLeuLeuTyrSerGluIleProLysValAlaGlyGlnAlaLeu 171
QY 311 CTGCTCTCTGCTGGAAACAGATGCTGATCTTCAGGCCACGCGCCACCATGGGCTC 370
Db 172 LeuLeuLeuSerTyrPylGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 191
QY 371 TACTCTCGAGAGAGAGACTGCTGAGGAGCGGAAAGCGCTGGGCTCGGCATCACG 430
Db 192 TyrSerAlaGlyGluGluLeuLeuArgGluAlaGlyAlaGlyLeuGlyValPheGlyIleThr 211
QY 431 TCCTACGACTTCCACAGGAGAGAGTGGCTCTCTCTTCCAGGCCAGACAGCTCTTC 490
Db 212 SerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 231
QY 491 CACTCTCCGAGCGCGGCGAAGACGCTTCATG-----GTGTC 529
Db 232 HisCysArgAspGlyGlyLysAsnGlyPheMetValSerProGlyProGlyCysValSer 251
QY 530 CCTATGAACCGCTGGAATCAAGACCCAGTGCCTGAGGCGCCCGGATGAGCCCAAAATC 589
Db 252 ProMetLysProLeuGlnIleLysThrGlnCysSerGlyProAlaGluMetAspProLysIle 271
QY 590 TGGCTGGCGGACCTGCTCTCTCTCTCAATTAACAGCAGCAGCTGGGTGGCGCAAC 649
Db 272 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrValAlaAsn 291
QY 650 ATCGAGACAGCGGAGAGCGCGGCTGACCTTCTGCGCAAGGTTTATCCAAATGCTCTG 709
Db 292 IleGluThrGlyGluGlnArgAlaGlyLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 311
QY 710 GATGACCCCAAGTGGCGGCTGCGCACCTCGCATACAGGAAGATTTGACCGCTTC 769

```

```
Dh 312 AspsapProlysserAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPhe 331
QY 770 ACTGGTACTGTTGGTGGCCACACGCTCTCGGAGAGTTTCAGAGGCGCTCAAGAGCTG 829
Dh 332 ThrIlyTrpTrpCysProThrAlaSerTrpLeu-----GluGlyLeuLysThrLeu 349
QY 830 CGAATCTGTATGAGGAGAGTGCATGAGTCCGAGGTGAGGTGCATTCACGTCCCTCTCT 889
Dh 350 ArgIleuTrpGluGluValAspGluSerGluValGluValIleHisValProSerPro 369
QY 890 GCGGTAGAGAAAGAGAGAGGAGCTGATGCGTATGCCAGACAGAGAGAGAAATGCC 949
Dh 370 AlaIeuGluGluArgLysThrAspSerTrpArgTrpProArgTrpGlySerLysAsnPro 389
QY 950 AAGATTCCTTGAAGACTGGCTGATGCCAGACTGACAGCCAGGAGCAATGCTCGAC 1009
Dh 390 LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 409
QY 1010 CAGGAGAAGAGCTGTGTCAGCCCTTCAGCTGCTGTTCCCGAAGTGAAGTACATGCC 1069
Dh 410 GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluTrpIleAla 429
QY 1070 AGGCGCGGG 1078
Dh 430 ArgAlaGly 432

RESULT 8
096NT8 PRELIMINARY; PRT; 312 AA.
ID 096NT8;
AC 096NT8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CNA FLJ30094 fis, clone BNGH4100034, weakly similar to dipeptidyl
DE peptidase IV (EC 3.4.14.5).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kikuchi H., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;
RT "MEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054656; BAB70784.1;
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9.1.
SQ SEQUENCE 312 AA; 35518 MW; ABE940AFC587717 CRC64;
```

Alignment Scores:

```
Pred. No.: 1,88e-103 Length: 312
Score: 1675.00 Matches: 311
Percent Similarity: 99.688 Conservative: 0
Best Local Similarity: 99.688 Mismatches: 1
Query Match: 34.93% Indels: 0
DB: 4 Gaps: 0
```

US-09-976-674-4 (1-2617) x 096NT8 (1-312)

```
QY 1664 ATGAGCCAGAACTTCGACATGTTGTCAGCCACTACAGACGCTGAGCAGCGCCCTGTC 1723
Dh 1 MetSerGlnAsnPheAspMetPheValSerHisTyrSerSerValSerThrProProCys 20
QY 1724 GTGACGCTACAAAGCGAGCGGCGCCGAGAGACGCCCTGCACAGACGCGCCGCTTC 1783
Dh 21 ValHisValTrpLysLeuSerGlyProAspAspProLeuHisLysGlnProArgPhe 40
```

```
QY 1784 TGCGGTAGCATGATGAGAGGAGCCAGCTGCCCGCCGATTAATGTTCCCTCAGAGATCTTC 1843
Dh 41 TrpAlaSerMetMetGluAlaAlaSerCysProProAspArgValAlaProProGluIlePhe 60
QY 1844 CATTTCCACACGCGCTCGATGTGCGGCTCTACGGCATGATTAACAAGCCACGCTTG 1903
Dh 61 HisPheHisThrArgSerAspValArgLeuTrpGlyMetIleTrpLysProHisAlaLeu 80
QY 1904 CAGCCAGGAGAAAGAGACCCACCGCTCTTTGATATGAGGCGCCCGAGTGCACCTG 1963
Dh 81 GlnProGlyLysLysHisProThrValLeuPheValTrpGlyGlyProGlnValGlnLeu 100
QY 1964 GTGAATACTCTTCGAAAGCATCAAGTACTTGGCTGCGCTCAACACACTGCTCCCTGGC 2023
Dh 101 ValAsnAsnSerPheLysGlyIleLysTrpLeuArgLeuAsnTrpIleAlaSerLeuGly 120
QY 2024 TACCGCTGTGTGATTGATGACGCGAGGGCTCTGTGACGAGGGCTTCGGTTGGAAGG 2083
Dh 121 TyrAlaValAlaValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGly 140
QY 2084 GCGCTGAAAAACCAATGGGCGAGGTGAGATGAGAGACGAGTGGAGGCGCTGCAGTTC 2143
Dh 141 AlaLeuLysAsnGlnMetGlyGlnValGlnIleGlnLysPheGlnValGluGlyLeuGlnPhe 160
QY 2144 GTGGCCGAGAAATGATGCTTCATGACCTGACCTGAGCCGAGTTCACATGCCTGCTCTAC 2203
Dh 161 ValAlaGluLysTrpGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTrp 180
QY 2204 GGGGGCTTCTCTGCTGTCATGAGGGCTAATCCACAAGCCCGGCTTCAAGTGGCCATC 2263
Dh 181 GlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 200
QY 2264 GCGGGCTCCCGCGTACCGTGTGGATGGCTAGCAGCAGGATACATGAGCGCTACATG 2323
Dh 201 AlaCysAlaProValTrpValTrpMetAlaArgTrpAspTrpGlyTrpGlnAlaGlyMet 220
QY 2324 GAGCTCCCTGAGAAACAACAGCAGCAGGCTATGAGGCGGGTCCGTGGCCCTGCAGCTGGAG 2383
Dh 221 AspValProGluLysAsnGlnHisGlyTrpGlnAlaGlySerValAlaLeuHisValGlu 240
QY 2384 AAGTGGCCCAATGAGCCCAACGCGTCTTATCTCTCAGCGCTTCTCGAGAAAGCTG 2443
Dh 241 LysLeuProAsnGlnProAsnArgLeuLeuIleHisGlyPheLeuAspGluAsnVal 260
QY 2444 CACTTTTCCACAAACTCTCTGCTCCCACTGATCCGAGAGGGAACCTTACAG 2503
Dh 261 HisPheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTrpGln 280
QY 2504 CTCGAGATCAACCCCAAGAGAGACAGATATCGCTGCCCGGAGTGGCGGAGCACTAT 2563
Dh 281 LeuGlnIleTrpProAsnGlnAlaGlnHisSerIleArgLysProGlnSerGlyGlnHisTyr 300
QY 2564 GAAGTCAAGTTCGCGCACTTTCACAGAAATACCTC 2599
Dh 301 GluValThrLeuLeuHisPheLeuGlnGluTrpLeu 312

RESULT 9
09VC20 PRELIMINARY; PRT; 1042 AA.
ID 09VC20;
AC 09VC20;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG3744 protein.
GN CG3744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
```


RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abdl J.F., Asgadyani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,
RA Ballew R.M., Baas A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck R.J., Brokstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA Hostal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazzo D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng H.O.,
RA Glabbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003749; JAI56357.1; --
DR MEROPS: S09_016; --
DR FlyBase: FBgn0039240; CG3744.
DR InterPro: IPR002469; DPrV_N_term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF009930; DPrV_N_term; 1.
DR Pfam: PF00326; Peptidase_S9; 1.
SQ SEQUENCE 1042 AA; 116706 MW; 5B915D5C365DF937 CRC64;

Alignment Scores:			
Pred. No.:	2,56e-100	Length:	1042
Score:	1631.50	Matches:	372
Percent Similarity:	54.76%	Conservative:	146
Best Local Similarity:	39.32%	Mismatches:	320
Query Match:	34.03%	Indels:	109
DB:	5	Gaps:	20
US-09-976-674-4 (1-2617) x Q9VC20 (1-1042)			
QY	5 CTTACACATGGCCACACCGGAGACCCCA-----ACGGCCGACCGAGGCGACGCA	52	
	: :		
Db	120 leuSerSerProthrSerGlyThrProProHisGlyLeuAspValAspGluGlyAsp	139	
QY	53 GCCGCGC-----ACAGATGACCCCGGCGC-----GCCCGC	79	
Db	140 GlucyAspAspCysGluGluAspGluAspAspProValAspAsnAspGluHisIleAla	159	
QY	80 TTCCAGGTCGACGAACCACTGCTGGGACGGGCTCCGGACATCATCCACGGCAGCCGAC	139	
	:		
Db	160 AlaProthrProAsnLysSerThrAlaGluAsnLysGlyIleValGluGluIleValGly	179	
QY	140 TACTCGGGCTCATTTGTCATCAACAGCGCGCCACGACATTCACAGTTTGGCAGAGACGAT	199	

Db	180	LysMetGlySerAsnLeuSerSerMetValProThrAsnValGlnPheArgHisLeuSerAsp	199
QY	200	GAGTCGTGGGCCCACTCCACCGCCTCTACTACCTGGGAATGCCATATGGCAGCCGAG	259
Db	200	GlyArgAla-----ArgCysTyrPheLeuGlyThrProProGlnSerThrGlu	215
QY	260	AACCCCTCCCTCTACTGTAGATTCCCAAGACAGTCGGGAAGAGGCTCTGCTGCT---	316
Db	216	ThrThrLeuLeuPheAlaAspIleAsnLeuThrGlnSerCysGlnGlnGlnLeuVal	235
QY	316	-----	316
Db	236	GlnArgLeuGlnGlyIleAlaSerAspGluTrpSerProThrMetAsnAlaGlySerPro	255
QY	317	-----CTGTCC	322
Db	256	ThrSerSerGlnHisGlnProAlaPheLeuPheAsnSerLeuProArgProArgLeuPro	275
QY	323	TGGAAAGCAGATGGCTGATCAT---TTCCAGGCCACACCCACCATTAGG-----	367
Db	276	TrpSerProLeuLeuGlnGlnProIleGlnSerSerGlyGlySerGlyGlySerGlySer	295
QY	368	-----GTCTACTCTCGGAGAGAGAGACTGCTTAGGAGGAGCGAAGACGCTGGGGCTTTC	421
Db	296	AlaSerProTyrAlaArgValGlyThrGlnLeuGlnGlnArgLysArgLeuSerThrTrp	315
QY	422	GGCATCACCTCTCAAGCTTCACACGAGAGTGGCCTTCTCTCTCCAGGCCACGAC	481
Db	316	GlyIleThrSerTyrGlnLeuHisLysProSerGlyLysLeuValPheProCysPheAsn	335
QY	482	AGCCCTTCCACTGTCCGACGGGGGCAAGACGGCTCATGGATGGTCCCTTCCGAACCG	541
Db	336	AspLeuTyrGlnCysLeuAspThrGlyTyrAsnSerGlyLeu-----LeuPhePro	352
QY	542	CTGGAAATCAAGACCCAGTCTCAGGGCCCGGATGAGCCCAAAATCTGCTCCGCGAC	601
Db	353	ThrGlnLeuArgThrCysProGlnTrpThrAlaLeuAspProGlnIleCysProGlnAsn	372
QY	602	CTGTGCTTCTTCTCTTCATCAATAACACGACGCTGGGTGGGCAACATCGACGACGC	661
Db	373	SerAspMetIleAlaTyrIleSerAspCysAspLeuPheValIleHisThrLeuSerGly	392
QY	662	GAGAGCGCGCGTACACTTCTGCGACCAAGATTATTCCAATGCTCTGGATACCCCAAG	721
Db	393	HisGlnLysArgLeuThrTyrThrSerThrGlyArgHisSerTyrValAspAspAlaLeu	412
QY	722	TCTCGGGGTGGGCCACTCTTCATACAGAGAAGATTTGACCCGCTCACTGGTACGCG	781
Db	413	SerAlaGlyValProSerTyrValMetGlnGlnIlePheSerArgTyrGlnIlePheTrp	432
QY	782	TGTGTCCCCACAGGCTCTGTGGGAAGGTTCAGAGGGCTCAAGACGCTGGAAATCTGTAT	841
Db	433	TrpGlnPro-----HisSerAsnAspGlyIle-----TyrArgIleValTyr	446
QY	842	GAGGAAGTCATGATGTCGAGGTGGAGATTCACAGTCGCTCTCTGCGCTAGAGAA	901
Db	447	GlnGlnValAspGlnSerGlnValSerValTyrThrPheProSerSerThrAlaMetHis	466
QY	902	AGGAAGAGCGACTGTATTCGGTACCCACAGACGAGGACAGAAATCCCAAGTATCCCTTG	961
Db	467	GlyArgValAspGluTyrArgPheProArgThrGlySerProAsnAlaLysSerLysLeu	486
QY	1022	CTGTGCGACGCCCTCACCCTGCGCTGGTTCGCCAGACGTGATACATGCCAGCGCGCGTGG	1081
Db	507	LeuProTyrSerLeuLeuAlaValPheSerTrpLeuGluTyrIleValArgValGlyTyr	526
QY	1082	ACCCGGATGGCAAAATACGCTTGGGACCATTTCTGTGACGGCGCCACAGAGGATGGCTCAG	1141

Dh 527 ThrProAspAlaLysTyrValTTrpValGInGlyLeuAspArgLysGInGlnArgLeuAsp 546
Qy 1142 CTCGTCTCTCCCTCCCGGCGCTGTTCATCCCGAGACAGAAATGAGACACGGCTA 1201
Dh 547 ValIleuLeuIleProLeuAspAsnPhcysGlnSerTyrSerSerGlnValSerThrPro 566
Qy 1202 GCGCTGTGCC-----AGAGCTGTCCCGAGAAATGTCACGCG-- 1237
Dh 567 ThrAspSerIleGlyAspAsnHisSerTrpArgSerLeuTyrSerArgTrpIleThrProLeu 586
Qy 1238 TATGTGGTGTACAGAGAGGAGTCACACAGCTGTGATGCATATGTTCATGACATTTTATCC 1257
Dh 587 GlnAlaIleuTyrThrGlnArgSerAspSerTrpIleAsnValHisAspMetLeuHisPhe 606
Qy 1298 TTCCCCCAATCAGAGAGGAGAGCAGCAGCTGTCTCCGCGCGCAATGAAATGCAAGAC 1357
Dh 607 LeuAspLeuThr-----GluThrSerValThrPheLeuThrAlaSerGlu---GluThr 623
Qy 1358 GCGTTGTGCCATTTTGTACAAAGTCACGCGC--GTTTAAATCCAGGCGCTACGATTTGG 1414
Dh 624 GlyPheArgHisLeuTyrLeuValThrAlaSerLeuLeuSerGlnAlaAsnGlyGln 643
Qy 1415 AGTAGGCCCTTCAGCCCGCGGGA-----GATGATTTTAAGTGC 1453
Dh 644 ProAspProGlySerValGlyAlaGlnProSerPheValAspLeuSerAlaLeuGlnPro 663
Qy 1454 CCCATTAGAGAGATGTGCTGTACACAGCGGTAGATGGAGTTTGGCGAGCGACGCGC 1513
Dh 664 ArgIleuAsnLysValAlaLeuThrSerGlyGlnTrpGlnValLeuAlaArgAsn--- 682
Qy 1514 TCCAGATCTGGGTCATGTAGAGAGACAGCTGGTGTACTTCCAGGCGACCAAGACAGC 1573
Dh 683 -----LeuTrpValAspLysAlaAsnLysLeuValTyrPheValGlyLeuArgAspThr 700
Qy 1574 CCGGTGAGAGACCCACTCTCAGTGTAGCTATGAGCGGCGCGGCGAGATGTACGGCTC 1633
Dh 701 ProLeuGlnLysHisLeuTyrValValSerLeuGlnArgProGlnHisIleArgLeuLeu 720
Qy 1634 ACCAGCGCGGCTTCCCATAGCTGTCTCATGTAGCCAGAACTTGCAGATGTCTGTACGC 1693
Dh 721 ThrGlnProGlyTyrSerTyrLeuValGlnPheAspArgLysLysLeuMetLeuLeu 740
Qy 1694 CACTACAGACGCTGAGACAGCGCGCGCTGCGTGCACGCTTACAGAGCTGAGCGCGCGGAC 1753
Dh 741 ValTyrLysAsnIleGln-ArgLeuProSerCysLys-----ValMetArgVal 756
Qy 1754 GAGCAGCCCGCTGACAGACAGCCCGGCTCTCGGCTGAC-----AT 1794
Dh 756 IAsnGlnThrCysSerAsnGlyValAsnGlyIleGlnIleSerLeuValGlyTyrLe 776
Qy 1795 GATGAGAGGACAGCGACGCGCGCGATTTATGTCTCTCAGAGATCTTCCATTTCCACAC 1854
Dh 776 UNHISGLIYGLYLPYSPROGILUPROGILNTRYCYS--ProGlnIlePheSerProGlnLe 795
Qy 1855 GCGGTGGATGTGGGCTCTACAGCATGATCTACAGCCCGCGCTTGCAGCGGCGGAA 1914
Dh 795 uProSerGlyAspIleValTyrAlaMetValPheLysProHisAsnPhcGluLeuGlyVal 815
Qy 1915 GAAGCAGCCCGCTCTCTTGTATGTAGAGCGCGCGAGTGTAGCGTGGATATATC 1974
Dh 815 LysTyrTrpThrValLeuAsnValTyrGlyGlyProGlnValGlnThrValAsnAsnTh 835
Qy 1975 CTTCAAAGGATCAAGTACTTGGCGCTCAACACAGCTGCGCTCCCTGCGCTACGCGCTGT 2034
Dh 835 rPheLysGlyLysHisGlnLeuArgMetHisMetLeuAlaIleGlnGlyTyrCysValI 855
Qy 2035 TGTGATTGACGGGCGGCTCTGTACAGCGAGGCGCTTGGTGTGACAGGCGCGTGA 2094
Dh 855 eCysIleAspSerArgLysArgLysArgGlyLysArgPheGlnSerHisIleArgG 875
Qy 2095 CCAATGGCGCAGGTGAGATTCGAGAGCGAGCGCTGAGGCGCTGAGTTCCGCGCGGAA 2154
Dh 875 yArgMetGlyGlnValGlnLeuThrAspGlnValAspAlaLeuArgSerLeuSerAspG 895

Qy 2155 GATAGCTTCATCGACCTGAGCGAGTTTGCCATTCATGCGTGGTCTTACGGGGCTTCT 2214
Dh 895 nLeuGlyTyrIleAspMetAspArgValAlaIleHisGlyTrpSerTyrGlyTyrLe 915
Qy 2215 CTCGCTCATGGGCTTAATCCACAAAGCCCGCAGGTCTTCAAGTGCGCATCCGGGTGCC 2274
Dh 915 uSerLeuMetGlyLeuValGlnTyrProLysIlePheLysValAlaIleAlaGlyAla 935
Qy 2275 GGTACCGGTGTGATGGCTACAGAGGATGCGGTCCGTGCGCGCTGAGAGAGTGCCTGA 2334
Dh 935 oValThrAsnTrpGlnTyrTrpAspTrpGlyTyrThrGlnArgTyrMetAspMetPro 955
Qy 2335 GAACACACACAGCGCTATGAGCGGCGGTCCGTGCGCGCTGAGAGAGTGCCTGA 2394
Dh 955 nAsnAsnGlnAlaGlyTyrSerAlaGlySerValLeuGlnTyrValAsnSerPhePro 975
Qy 2395 TGAAGCCCAACCGCTTGTATTCCTTCACAGCGCTCTCTGAGCGAAACCTGACATTTTCA 2454
Dh 975 uGlnAspLysArgLeuLeuLeuIleHisGlyLeuIleAspGlnAsnValHisPheCysH 995
Qy 2455 CACAACTCTCTGCTCCCACTGATCCGAGCGAGGAAACCTTACACAGCTGACATCTA 2514
Dh 995 sThrSerArgLeuIleSerAlaLeuAsnLysAlaAsnLysArgProTyrGlnValHisLeu 1015
Qy 2515 CCCCAACAGAGACAGATATTCGCTCCCGAGTCCGCGGCGACATGTGAATCACGTT 2574
Dh 1015 eProGlnGlnArgHisSerLeuArgAsnLeuGlnLysSerAsnLysAsnTyrGlnTrpLys 1035
Qy 2575 GCTGCACTTCTCTACAG 2590
Dh 1035 uLeuSerPheLeuGln 1040
RESULT 10
ID 09VC19 PRELIMINARY: PRT: 1102 AA.
AC 09VC19;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE G33744 protein.
GN G33744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amandlides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Man K.H., Doyle C., Baxter E.G., Blazek R.G., Champe M., Pfeiffer B.D.,
RA April J.F., Abdayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballou R.M., Besu A.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,


```

Db 704 ProAspProGlySerValGlyAlaIleProSerPheValAlaSerLeuSerAlaLeuGlnPro 723
QY 1454 CCCATTAAAGAGAGATTGCTGTGACAGCGGTGAATGGAGGTTTGGCCAGGACGGC 1513
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 724 ArgIleuAsnLysValAlaIleuThrSerGlyIuTrpGlnValIleuAlaArgsn--- 742
QY 1514 TCCAAAGATCTGGGTCAATGAGAGACCAAGCTGGTGTACTTCCAGGGACCAAGACAGC 1573
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 743 -----LeuTrpValAlaSprLysAlaSnLysLeuValTyrrPheValGlyLeuArgSprThr 760
QY 1574 CCGGTGAGACGACAGCTGTAGGTGATGAGGGCGGCGGCGGAGATGTACGGCTC 1633
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 761 ProLeuGlnLysHisLeuTyrrValAlaSerLeuGlnLysArgProGlnHisIleArgLeuLeu 780
QY 1634 ACCAGCGCGGCTTCTCCATGATGCTGTCCATGAGCCAGAACTTGCATGTTCCATGAC 1693
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 ThrGluProGlyTyrrSerTyrrLeuValGluPheAspArgLysCysLysLeuMetLeuLeu 800
QY 1694 CACTACAGACGCGGTGACAGCGCGGCTGCGTGCACGTTCTACAGCTGACGGCGCCGAC 1753
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 801 ValTyrrCysAsnIleGln-ArgLeuProSerCysLys-----ValMetArgVa 816
QY 1754 GACGACCCCTGACACAGACGCGCGCTTCGCGCTAGC-----AT 1794
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 816 LAsnGlnThrCysSerAsnGlyValAlaAsnGlyIleGlnIleSerLeuValGlyTyrrLe 836
QY 1795 GATGAGAGACGACGACGCTCCCGGATTAATGTTCTCCAGATCTTCCATTTCCACAC 1854
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 836 uHisGlnGlyLysProGlnIuTrpCys---ProGlnIlePheSerProGlnIle 855
QY 1855 GCGCTGGATGTGCGGCTCTACGCGCATATACAGCCCGACGCGTTCAGCGGAGAA 1914
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 855 uProSerGlyAspIleValTyrrAlaMetValPheLysProHisAsnPheGlnLeuGlyVa 875
QY 1915 GAAGACCCCGCCGCGCTTGTATATGAGGGCCCGAGGCGCGGCGGAGAAATATC 1974
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 875 LysTyrrProThrValIleuAsnValTyrrGlyProGlnValGlnThrValAsnAsnTh 895
QY 1975 CTTCAAAGGCAATGAATCTTGGCGGCTCAACACACTGCGCTCCGCGGTACGCGGTG 2034
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 895 rPheLysGlyLysHisGlnLeuArgMetHisMetLeuAlaGlnGlyTyrrCysValIl 915
QY 2035 TGTGATGACGCGGAGGCGCTCTGACGAGAGGCTTGGTTGCAAGGCGCGCTGA 2094
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 915 eCysIleAspSerArgGlySerArgIleArgGlyLysArgPheGlnSerHisIleArgGl 935
QY 2095 CCAATGGCGGAGGTGAGATGAGAGGAGGAGGCGCTGCGATTCGTGCGCGAGAA 2154
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 935 YArgMetGlyIuValGlyLeuThrAspGlnValAlaSerAlaLeuArgSerLeuSerAspGl 955
QY 2155 GTATGGCTTCAATCGACGCGGAGTGGCCATGCGGTGCTGACGGGCGCTTCT 2214
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 nLeuGlyIuTrIleAspMetLAspArgValAlaIleHisGlyIuTrpSerTyrrGlyTyrrLe 975
QY 2215 CTGCGTATGGGCGCTAATCCAAAGCCCGAGGTGTTCAAGTGCGCATCGCGGCTGCC 2274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 975 uSerLeuMetGlyLeuValGlnTyrrProLysIlePheLysValAlaIleAlaGlyAlaPr 995
QY 2275 GGTGACGCTTGGATGGCGCTGACGACAGGATGACAGGCGCTGACATGGAGCGTCCCG 2334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 995 oValThrAsnTrpGluTyrrTyrrAspThrGlyTyrrThrIuArgTyrrMetLAspMetProGl 1015
QY 2335 GAACACGACGACGCTATGAGGCGGAGTTCGTGCGCGCTGACGCTGAGAGAGCTGCCAA 2394
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1015 nAsnAsnGlnLysGlyTyrrSerAlaGlySerValLeuGlnLysValAlaAsnSerPheProGl 1035
QY 2395 TGAGCCCAACGCGCTTATATCTCCACGCGCTTCCGAGCAAAAGCTGCACTTTTCC 2454
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1035 uGlnAspLysArgLeuLeuLeuIleHisGlyLeuIleAspLysValAlaHisPheCysHis 1055
QY 2455 CACAACCTCTCGTCCCAAGTATGAGCGGAGGAGGAGAACTTACAGCTCCAGATGTA 2514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1055 sThrSerArgLeuIleSerAlaLeuAsnLysAlaAsnLysProTyrrGlnValHisLeuPh 1075

```

```

QY 2515 CCCAAGCAGAGACAGATATTCGCTGCCCGGAGTGGGCGGAGCACTATGAGTACGTT 2574
    :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 eProGlnGluArgHisSerLeuArgAsnLeuGlnSerHisLysAsnTyrrGluThrLysLe 1095
QY 2575 GCTGCATTTTCTACAG 2590
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1095 uLeuSerPheLeuGln 1100

RESULT 11
Q9HBM3 PRELIMINARY: PRT: 465 AA.
AC Q9HBM3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolfitt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221636; AAG29768.1; -.
DR MEROPS; S09.018; -.
FT NON_TER
FT 1
SQ SEQUENCE 465 AA; 53197 MW; 22939BC0A4FE29CE CRC64;

Alignment Scores:
Pred. No.: 2,47e-91 Length: 465
Score: 1494.50 Matches: 281
Percent Similarity: 62.95% Conservative: 69
Best Local Similarity: 50.54% Mismatches: 105
Query Match: 31.17% Indels: 101
Gaps: 2

US-09-976-674-4 (1-2617) x Q9HBM3 (1-465)
QY 932 ACAGGACAGCAAGATCCCAAGATGCTTGAACATGCTGAGTTCACATGACAGCCAG 991
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ThrGlyThrAlaAsnProLysValThrPheLysMetSerGluIleMetIleAspAlaGlu 20
QY 992 GCAAGATCGTCTGACCCAGGAGAGAGAGGCTGGTGCAGCGCTTCAGCTCGTGTCCG 1051
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 21 GlyArgIleIleAspValIleAspLysGlnLeuIleGlnProPheGlnIleLeuPheLeu 40
QY 1052 AAGGTGAGTACATGCCAGGCGGCGGTGACCCCGATGGCAATATACGCTGGGCCATG 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 GlyValGlyTyrrIleAlaArgAlaGlyThrPheProGlnGlyLysTyrrAlaTrpSerIle 60
QY 1112 TTCTGGACCGGCGCCAGACATGCTGCTGACGCTGCTCCCTCCCGGCGGCTGTATC 1171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 LeuLeuAspArgSerGlnThrArgLeuGlnIleValLeuIleSerProGlnLeuPheIle 80
QY 1172 CCGAGCAGAGAAATGAGAGGACGAGCGCTACGCTCTGCAAGAGTGTCCCGAGAAATGC 1231
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 81 ProValGlnAspArgValMetGlnArgGlnArgLeuIleGlnSerValTrpAspSerVal 100
QY 1232 CAGCGTATGTGGTGTACGAGAGAGTCCAAAGCTGTGATCAATGTTCATGACATCTTC 1291
    ||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 101 ThrProLeuIleIleTyrrGlnGluThrThrAspIleTrpIleAsnIleHisAspIlePhe 120
QY 1292 TATCCCTTCCCAATGACAGAGGAGAGAGACAGCTGTCTTCCCGGCGCAATGATGC 1351
    :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 HisValPheProGlnSerHis--GluGlnIuIleGlnPheIlePheAlaSerGlnCys 139

```

QY 1352 AAGACCGGCTTCTGCACTTTGTAACAAGTACCGCGGCTTTAAATCCAGGCTACGAT 1411
 Db 140 LysThrGlyPheAlaGlnHisLeuTyrIleThrSerIleLeuLysGlnSerLysTyrLys 159
 QY 1412 TGGAGTAGCCCTTACGCCCCGGGGAGATGAATTAATGATGCCCATTAAGAGAGAT 1471
 Db 160 ArgSerSerGlyGlyLeuProAlaProSerAspPheLysCysProIleLysGlnGluIle 179
 QY 1472 GCTCGACACCGCGTGAATGGAGTTTGGCGAGCGACCGGCTCCAAATATCGGGTCAAT 1531
 Db 180 AlaIleThrSerGlyGlyTyrPheIleuValLeuGlyArgHisGlySerAsnIleGlnValAsp 199
 QY 1532 GAGGAGACCAAGCTGACTACTACAGCTACAGGCGACACGAGCGCGCTGGAGACCAACCTC 1591
 Db 200 GluValAlaGlyLeuValTyrPheGlnGlyThrLysAspSerProLeuGlnHisIleu 219
 QY 1592 TACGTGTACGATATGAGCGCGCGAGATGTACGCTCACACGCGCGCTTCTCC 1651
 Db 220 TyrValValSerTyrValAsnProGlyGlnValThrArgLeuThrAspArgGlyTyrSer 239
 QY 1652 CATAGCTCTGCATGACCCAGACACTTCACATGTTCTGCAGCCACTACAGACGGTGAGC 1711
 Db 240 HisSerCysCysIleSerGlnHisCysAspPheIleSerLysTyrSerAsnGlnLys 259
 QY 1712 ACGCGCGCTGCGTACAGCTACAGCTACAGGCGCGCGCGCGCGCGCTGACACAG 1771
 Db 260 AsnProHisCysValSerLeuTyrLysLeuSerSerProGluAspAspProThrCysLys 279
 QY 1772 CAGCCCCCTCTGCGTAGCATGATGAGAGCGACCGACCTGCCCGCGGATATGTTCT 1831
 Db 280 ThrLysGlnPheTyrAlaThrIleLeuAspSerAlaGlyProLeuProAspTyrThrPro 299
 QY 1832 CCAGAGATCTTCGATTCACACGCGCTCGGATGTGCGGCTACGCGCATGTCTACAG 1891
 Db 300 ProGluIlePheSerPheGlnSerThrThrGlyPheThrLysGlyMetLeuTyrLys 319
 QY 1892 CCCCAGCGCTTCTGCGTAGCATGATGAGAGCGACCGCGCTGTTGTTATGAGAGCGCC 1951
 Db 320 ProHisAspLeuGlnProGlyLysLysTyrProThrValLeuPheIleTyrGlyLysPro 339
 QY 1952 CAGGTGACGTGTGTAATTAATCTCTTCAAGGCGATCAACTGCGGCTCAACACACTG 2011
 Db 340 Gln----- 340
 QY 2012 GCCTTCCTGGGCTACGCGGTGGTGTGATGACGCGAGGGGCTCTGTCAAGGAGGCT 2071
 Db 340 ----- 340
 QY 2072 CGGTTCGAAGGGGCGCTGAAAAACAATGGGCGCAGGTGGAGATGAGAGACCGAGTGAG 2131
 Db 340 ----- 340
 QY 2132 GGCGTCGATTCGTGCGCGAGAGATGATGGCTTCATGCACTGAGCGAGTTCATTCAT 2191
 Db 340 ----- 340
 QY 2192 GGCTGGTCTTACGGGGGCTTCTCTCGCTCATGGGCTAATCCACAAGCCGAGGTTC 2251
 Db 340 ----- 340
 QY 2252 AAGGTGGCATCGGGGGTCCCGGTCACCGCTGTGATGGCTAGACACAGGATACACT 2311
 Db 341 ---ValAlaIleAlaGlyAlaProValThrLeuThrPheThrLysAspThrGlyTyrThr 359
 QY 2312 GAGCGCTACATGAGAGTCCCTGAGAACACACGACGCGCTATGAGCGGGTTCCTGTGCC 2371
 Db 360 GluArgTyrMetGlyHisProAspGlnAsnGlnGlnGlyTyrTyrLysGlnLysValAla 379
 QY 2372 CTGACGCTGGAGAGAGCTGGCCATGAGCCCAACGCGTGTGCTTATCTCCAGCGTTCCTG 2431
 Db 380 MetGlnAlaGlnLysPheProSerGlnProAsnArgLeuLeuLeuHisGlyLysPheLeu 399

QY 2432 GACGAAAAAGTGCACACTTTTCCACAAACTCTCGTCTCCCAACTGATCGAGCAGG 2491
 Db 400 AspGluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValAlaArgAlaGly 419
 QY 2492 AAACCTTACCACTCCAGATCCAGATTAACCCCAAGAGACACAGATTTCCGTCGCCAGTGG 2551
 Db 420 LysProTyrAspLeuGlnIleTyrProGlnGlnAlaGlnHisSerIleArgValProGluSer 439
 QY 2552 GCGGAGCACTATGAGTCACTGCGTGCATCTTTCACAGGAATACCTC 2599
 Db 440 GlyGlnHisTyrGlnLeuHisLeuLeuHisTyrLeuGlnGlnAsnLeu 455
 RESULT 12
 Q9NMF4 PRELIMINARY: PRT: 469 AA.
 AC Q9NMF4
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CDNA FLJ20283 fis, clone HEP04088.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Oabayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isoigal T., Sugano S.;
 RT "NEBO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000290; BAA91059.1; -
 DR MEROPS: S09.018; -
 SQ SEQUENCE 469 AA: 54367 MW: 088ED0B1E46C1F CRC64;
 Alignment Scores:
 Pred. No.: 2e-83 Length: 469
 Score: 1376.00 Matches: 245
 Percent Similarity: 75.51% Conservative: 85
 Best Local Similarity: 56.06% Mismatches: 105
 Query Match: 28.70% Indels: 2
 DB: 4 Gaps: 2
 US-09-976-674-4 (1-2617) x Q9NMF4 (1-469)
 QY 80 TTCAGGTGCAGAGCACTGTGGAGCGGCTCGAGACATCAACGCGAGCGCAAG 139
 Db 35 PheTyrValGlnArgTyrSerThrPheGlnLysLysLeuAlaAspThrArgLys 54
 QY 140 TACTGGGCGCTATGTCAACAAGCGCGCCAGCACTTCGATTTGTGCAAGAAGCGGAT 199
 Db 55 TyrHisGlyTyrMetLeuAlaLysAlaProHisAspPheMetPheValLysArgAsnAsp 74
 QY 200 GAGTGTGGGCGCCAGTCCACCGCGCTCTACTCTGAGGAATGCAATATGCGAGCGGAGAG 259
 Db 75 ProAspGlyProHisSerAspArgIleTyrTyrLeuAlaMetSerGlyGlnAsnArgGlu 94
 QY 260 AACTCCCTCTACTCTGATGATCCCAAGAAGTCCGGAAGAGAGGCTCTGCTCTG 319
 Db 95 AsnThrLeuPheTyrSerGlnIleProLysThrIleAsnArgAlaAlaValLeuMetLeu 114
 QY 320 TCCGTGAAGCAATGCTGATATTCACAGGCGCGCGCCACCAATGCGGTCTACTCTGG 379
 Db 115 SerTyrLysProLeuAsnAspLeuPheGlnAlaThrLeuAspTyrGlyMetTyrSerArg 134
 QY 380 GAGGAGAGCTCTGAGGAGCGGAAACGCTGGGGGCTTCTGGCATCAGCTCAAGAC 439
 Db 135 GlnGlnGlnLeuLeuArgGlnArgLysArgIleGlyThrValGlyIleLeuAspTyrAsp 154
 QY 440 TTCACAGCGAGATGGCGCTTCTCTTCCAGGCGAGACAGCGCTTCCAGCTGTGCG 499
 Db 155 TyrHisGlnGlySerGlyThrPheLeuPheGlnAlaGlySerGlyIleTyrHisValLys 174

```

OY 500 GAGCGCGGCAAGAAAGGCTTCATGGTGTCCTCCATGAACCGCTGGAAATCAAGACCCAG 559
    |||||
    |||||
    |||||
Db 175 AspIlglygProglnglypHeThrglngInlProleuAtrpArosluValGluIurhSer 194
    |||||
    |||||
    |||||
OY 560 TGCACAGGCGCCCGGATGGACCCCAAAATCTGCCCTGGCCGCCCTCTCTCCCTTC 619
    |||||
    |||||
    |||||
Db 195 CysProasnlleAqMetAspProlysluEncysProAlaAspProAspTrpIleAlaHe 214
    |||||
    |||||
    |||||
OY 620 ATCAATTAACACGACCTCTGGGTGGCCAACTCGACAGGCGAGGACGGCGGCTGACC 679
    |||||
    |||||
    |||||
Db 215 lHeHsSerAsnAspIleTrpIleSerAsnIleValThraAgIugluIurArgIeuhr 234
    |||||
    |||||
    |||||
OY 680 TTCTGCCACCAAGGTTTATTCATATGCTCGATGACCCCAAGTCTCGGGTGGCCACC 739
    |||||
    |||||
    |||||
Db 235 TyrValAlaAsnGluIeuAlaAsnMetGluIuAspAlaArgSerAlaGluValAlaThr 254
    |||||
    |||||
    |||||
OY 740 TTGCTCATACAGGAAGGTTGACCGGCTTCACCTGGGTCTGGTGGTGGCCACAGCCCTCC 799
    |||||
    |||||
    |||||
Db 255 PheValIeuGluIugluIuPheAspArgTyrSerGlyTyrTrpPcysProIuAspIaGlu 274
    |||||
    |||||
    |||||
OY 800 TGGGAAGGTTTCAGAGGCGCTTCAGACGCTCGAAATCTGTATGAGAAGTGCATGACTCC 859
    |||||
    |||||
    |||||
Db 275 ThrThrProSerGlyGly---LysIleuAurGlyIeuTyrGluIuAsnAspGluSer 293
    |||||
    |||||
    |||||
OY 860 GAGGTGAGGTCAATTCACGTCCTCTCTCGCTAGAGAAGAAGACGCGACTCGTAT 919
    |||||
    |||||
    |||||
Db 294 GluValAlaGluIleIleHsValThrSerProMetIeuGluIurThrArgAlaAspSerPhe 313
    |||||
    |||||
    |||||
OY 920 CGGTACCCCAAGACAGACAGAAATCCCAAGATGCTGCTGAAGTGGCTGAGTTCAG 979
    |||||
    |||||
    |||||
Db 314 ArgTyrProIuSlyThrGlyThAlaAsnProIuSvalThrPheIuSMeSerGluIleMet 333
    |||||
    |||||
    |||||
OY 980 ACTGACACCCAGGCGCAAGATGCTTCGACCCAGAGAAGAAGAGTGTGTCACGCTTCAGC 1039
    |||||
    |||||
    |||||
Db 334 lIeAspAlaGluIuArgIleIleAspValIleAspIuSgluIeuIleGlnProPheGlu 353
    |||||
    |||||
    |||||
OY 1040 TCCTGTTCCTCCGAAGGTGAGTACATGCCACAGGCGGGTGGACCCGGATGCAAAATAC 1099
    |||||
    |||||
    |||||
Db 354 lIeAurHeuIeGluIuGlyAlaGlyTyrIleAlaArgAlaGlyTyrProGluGlyLysTyr 373
    |||||
    |||||
    |||||
OY 1100 GCGTGGGCGCATGTCCTGGACCGGCGCCAGCATGCTGCTCCAGCTCTCCCTCCCGG 1159
    |||||
    |||||
    |||||
Db 374 AlaTrpSerIleIeuLeuAspArgSerGlnThrArgIeuGluIleValIeuIleSerPro 393
    |||||
    |||||
    |||||
OY 1160 GCGCTGTTCATCCGACACACAGCAATGAGGACGCGGCTAGCGCTTCGCCAGAGCTGTC 1219
    |||||
    |||||
    |||||
Db 394 GluAurHeuIeIleProValAlaAspAspValMetGluAurGluIuArgIeuIleGluSerVal 413
    |||||
    |||||
    |||||
OY 1220 CCCAGGAATGTCACCGCTATGTGTGTCAGAGAGAGGTCAACCAAGTCTGATCAATGT 1279
    |||||
    |||||
    |||||
Db 414 ProAspSerValThrProIeuIleIleTyrGluIuThrThrAspIleTrpIleAsnIle 433
    |||||
    |||||
    |||||
OY 1280 CATGACATCTCTATCCCTCCCTCCCAATCAGAGGAGAGAGAGAGCTGCTTCTCGCC 1339
    |||||
    |||||
    |||||
Db 434 HisAspIlePheHsValPheProGlnSerHis---GluIuGluIuIleGluPheIlePhe 452
    |||||
    |||||
    |||||
OY 1340 GCCAATGATGACAGACCGGCTTCGCCATTTGTACAAAGTCAACGCGCGT 1390
    |||||
    |||||
    |||||
Db 453 AlAserGluCysIuSlyThrGlyPheArgHisIeuTyrIuSlyIleThrSerIle 469
    |||||
    |||||
    |||||

```

RESULT 13

```

OY 09HBM2 PRELIMINARY; PRT; 360 AA.
AC 09HBM2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Dipeptidyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBL_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; Pubmed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
RA Correll M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL: AF221637; AAC29769.1; -.
DR MEROPS: S09.018; -.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 360 AA; 41070 MW; CF81C0BB61423E22 CRC64;

```

Alignment Scores:

```

Pred. No.: 4,21e-73 Length: 360
Score: 1220.50 Matches: 236
Percent Similarity: 54.31% Conservative: 54
Best local Similarity: 44.19% Mismatches: 69
Query Match: 25.45% Indels: 175
DB: Gaps: 2

```

US-09-976-674-4 (1-2617) x 09HBM2 (1-360)

```

OY 710 GATGACCCCAAGTCTGGCGGTGTGGCCACCTTCGTCATACAGAAAGTTCGACCGCTTC 769
    |||||
    |||||
    |||||
Db 2 GluAspAlaAspSerAlaGlyValAlaThrPheValIeuGluIuPheAspArgTyr 21
    |||||
    |||||
    |||||
OY 770 ACTGGGTACTGTGGTGGCCACAGCCTCTCGGGAAGGTTCAGAGGCGCTCAACAGCTG 829
    |||||
    |||||
    |||||
Db 22 SerGlyTyrTrpPcysProIuSvalGluIurThrProSerGlyGly---LysIleu 40
    |||||
    |||||
    |||||
OY 830 CGAATCTGTATGAGAAAGTGTGATGAGTCCGAGGTGAGGTCAATTCACGTCCTCTCT 889
    |||||
    |||||
    |||||
Db 41 ArgIleuTyrGluIuAsnAspGluSerGluValAlaGluIleIleHsValThrSerPro 60
    |||||
    |||||
    |||||
OY 890 GCGGTGAGAAGAAGAGAGGACTCGTATCGTATGCCACGACAGGACGCAAGATCCC 949
    |||||
    |||||
    |||||
Db 61 MetIeuGluIurThrArgAlaAspSerPheArgTyrProIuSlyThrAlaAsnPro 80
    |||||
    |||||
    |||||
OY 950 AAGATTGCTTGAAGAACTGCTGAGTTCAGACTGACAGACCCAGGCAAGATGCTTCGACC 1009
    |||||
    |||||
    |||||
Db 81 LysValThrPheIuSMeSerGluIleMetIleAspAlaGluIuArgIleIle 98
    |||||
    |||||
    |||||
OY 1010 CAGAGAAAGAGCTGTGTGCACCGCTTCAGCTCGCTGTTCCGAAGTGGAGTACATGCC 1069
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1070 AGGCGCGGTGACCCCGGATGGCAAAATACGCTGGGCCATGTTCTGGACCGGCCAG 1129
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1130 CAGTGGCTCAGCTGTCCTCTCCCGCGGCGCTGTTCATCCGAGACAGAGAAATGAG 1189
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1190 GAGCAGCGGCTAGGCTCTGCCAGAGCTGTCCCAAGAAATGTCAGCCGTATGTGTGATC 1249
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1250 GAGGAGGTCAACCAAGTGTGATGATCATATGTTTCATCATATCTTATCCCTTCCCAATCA 1309
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1310 GAGGAGAGACAGAGCTGTCTTCTCCGCGCAATGAAATGACAGACCGGCTTGTGCAT 1369
    |||||
    |||||
    |||||
Db 98 ----- 98
OY 1370 TTGTACAAAGTCAACGCGCTTTTAAATCCAGAGGCTACGATTTGGAGTGAAGCTTCAGC 1429
    |||||
    |||||
    |||||

```

```

Db 98 ----- 98
OY 1430 CCGGGGGAAGATGATTTAAGTCCCAATTAAAGAAAGATTGCTCTGACACAGCGGTGA 1489
Db 98 ----- 98
OY 1490 TGGAGGTTTGGCGAGCGACGGCTCCAAAGATCTGGTCAATGAGAGACCAAGCTGGT 1549
Db 99 -----ValasgIuValAArgAtgIeuVal 106
OY 1550 TACTTCAGAGGACCAAGACAGCGCGTGGAGACACACCTTACGTGGTGCATATGAG 1609
Db 107 TyrrheglugIyThrllysAspserProleuGlnHshIsleuTyValValSerTyVal 126
OY 1610 GCGGGCGGAGATGATGAGCTCACACAGCGCGCTTCCCATAGCTGCTCCATGAGC 1669
Db 127 AsnProglIyGluValThrlAArgLeuThrsParArgIyTySerHshIsSerTyScysIleSer 146
OY 1670 CAGAACTTCGACATGTTGTCGACACACAGCGGTGACAGCGCGCGCTGGCGAC 1729
Db 147 GlnHshCysAspRphrheIleSerIyTySerAsnGlnIyAsnProHshCysValSer 166
OY 1730 GTCTACAAAGCTGAGGGCGCCGACAGCAGACCCCTGCACAMAGCAGCCCGCTTGGGCT 1789
Db 167 LeuTyrrIleuSerSerProglIuAspAspProthrcCysIyThrlYsgIuIubhertpala 186
OY 1790 AGCATGATGGAGGACGACAGCTGCCCCGATATGTTCTCCAGAGATCTTCATTTG 1849
Db 187 ThrIleuAspSerAlaGlyProleuProAspTyrrhProProglIuIlePheSerPhe 206
OY 1850 CACAGCGCTCGATGTCGCGCTACGCGATGATCTACAGCCCGACCGCTTGCAGCA 1909
Db 207 GluSerThrlhrlGlyPheThrlleuTyrrIyGlyMeuIeTyrrIyLysProHshIsleuGlnPro 226
OY 1910 GGGAGAAAGCACCCGACCGCTCTTGTATATGAGAGCGCCGACGTCAGCTGTGAAT 1969
Db 227 GlyIyLysTyrrProthValleuPheIleTyrrGlyIyProglIuValGlnIleuValAsn 246
OY 1970 AACTCTTCAAAGGATCATAGTACTGGGGCTCAACACACCTGGCTCCCTGGGCTACGCC 2029
Db 247 AsnaIyRphrlYsgIyValIyLysTyrrPheArTyrrIeuaInThrlleuAlaSerIeuIyTyrrVal 266
OY 2030 GTGGTGTGATGACGCGAGCGGGCTCTGTGACGCGAGGCTTCGGGTTCGAAGGGCGCTG 2089
Db 267 ValIyValIleAspAsnArgIySerCysHshArgGlyLeuIyShheIuIubIleAlaPhe 286
OY 2090 AAAAAACAATGGGCGAGGTGAGATCGAGACAGCAGGTGAGGGCGCTGCAGTTCGTGGCC 2149
Db 287 LysTyrrIyMetGlyGlnIleGlnIleAspAspGlnValGlnIyLeuIuIyTyrrAla 306
OY 2150 GAGAACTATGGCTTATGACCTGAGCCGAGTTCGATTCATGAGCTGCTTACGGGGGCT 2209
Db 307 SerArTyrrAspRphrlleAspRleuAspArgValIyGlyIleHshIyTyrrPheTyrrGlyIy 326
OY 2210 TTCCTCGCTGATGGGGTATATCCACAAGCCGAGGTGTCGAAGTGGCCATTCGGGGCT 2269
Db 327 TyrrIeSerIeuMetAlaIeuMetGlnArgSerAspIlePheArgValAlaIleAlaGly 346
OY 2270 GCCCGGTCACCGCTGTGATGGCTACGACACAGGGGTACACT 2311
Db 347 AlaProValThrlleuTrIlePheTyrrAspThrlGlyTyrr 360

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woolliatt E., Sutherland G.R., McCaughan G.W.,
  Correll M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
  dipeptidyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL: AF221635; AAC29767.1; -.
DR MEROPS: S09.018; -.
DR InterPro: IPR001375; Peptidase_S9.
DR Pfam: PF00326; Peptidase_S9; 1.
FT NON_TER 1
SQ SEQUENCE 310 AA; 35396 MW; 8E87C34026D9C7AC CRC64;

Alignment Scores:
Pred. No.: 1,17e-64 Length: 310
Score: 1093.50 Matches: 200
Percent Similarity: 69.34% Conservative: 42
Best Local Similarity: 57.31% Mismatches: 58
Query Match: 22.81% Indels: 49
DB: Gaps: 1

US-09-976-674-4 (1-2617) x Q9HBM4 (1-310)
OY 1553 TTCAGGGGACCAAGAGACAGCGCGTGGAGACACACCTTACGTGGTGCATATGAGCG 1612
Db 1 PheGlnGlyThrllysAspserProleuGlnHshIsleuTyValValSerTyValAsn 20
OY 1613 GCGGGGAGATGTCAGCTGCTCACACAGCGCGCTTCCCATAGCTGCTCCATGAGCCAG 1672
Db 21 ProglIyValIleThrlAArgLeuThrsParArgIyTySerHshIsSerTyScysIleSerGln 40
OY 1673 AACTTCGACATGTCGTCGACACACAGCGGTGACAGCGCGCGCTGGCGACGTC 1732
Db 41 HshCysAspRphrheIleSerIyTySerAsnGlnIyAsnProHshCysValSerIeu 60
OY 1733 TACAGCTGAGCGCGCCGACAGCAGACCCCTGCACAMAGCAGCCCGCTTGGGCTAGC 1792
Db 61 TyrrIyLseuSerSerProglIuAspAspProthrcCysIyThrlYsgIuIubhertpalaThr 80
OY 1793 ATGATGAGGACGACAGCTGCCCCGAGTATGTTCTTCCAGAGATCTTCATTTGCAC 1852
Db 81 IleuAspSerAlaGlyProleuProAspTyrrhProProglIuIlePheSerPheGln 100
OY 1853 ACGCGCTCGAGTGTGGGCTGTACGCGATGATCTTCCAGAGATCTTCATTTGCAC 1912
Db 101 SerThrlhrlGlyPheThrlleuTyrrIyGlyMeuIeTyrrIyLysProHshIsleuGlnProGly 120
OY 1913 AAGAAACACCCGACCGCTCTTGTATATGAGAGCGCCGAGTGCAGCTGGTGAATAAC 1972
Db 121 LysIyTyrrProthValleuPheIleTyrrGlyIyProglIn----- 134
OY 1973 TCCCTCAAGGATCAAGTACTTGGCGGTCAACACAGCTGCCCTCCCTGGGCTACGCCGTG 2032
Db 134 ----- 134
OY 2033 GTTGTGATTGAGGCGAGGGGCTCTGTGACGAGGCTTCGGTTCGAAGGGCGCTGAAA 2092
Db 134 ----- 134
OY 2093 AACCAATGGGCGAGGTGAGATCGAGACACAGGTGAGGGCGCTGCAGTTGCGCCGAG 2152
Db 135 -----GlyGlnIleIuIleAspAspGlnValGlnIyLeuIuIyTyrrAlaIleAsp 151
OY 2153 AAGTATGCTTATGACCTGAGCCGAGTTCGATTCATGAGCTGCTTACGGGGGCTTC 2212
Db 152 ArgTyrrAspRphrlleAspRleuAspArgValIyGlyIleHshIyTyrrPheTyrrGlyIyTy 171
OY 2213 CTTCTGCTCATGAGGGGTATATCCACAAGCCGAGGTTCGAAGTGGCCATTCGGGGGTGCC 2272

```

Db	172	LeuSerLeuMetAlaLeuMetGlnArgSerAspIlePheArgValAlaIleAlaGlyAla	191
QY	2273	CCGGTCACCCGCTTGATGGCCATGACACACAGGGTACATGAGCGGTACGCTTC	233
Db	192	ProValThrLeuTrpIlePheTyrAspPheGlyTyrThrGlnArgTyrMetGlyHisPro	211
QY	2333	GAGAAACACCGACGCGGTATGAGCGGGGTTCGGTGGCCCTGCACGTCGGAGAACTCGCC	239
Db	212	AspGlnAsnGlnGlnGlyTyrTyrLeuGlySerValAlaMetGlnAlaGlyIlePhePro	231
QY	2393	AATGAGCCCAACCGCTTGCTTATCTCCACGCGCTTCCTGGACGAAACGTCGACTTTTC	245
Db	232	SerGlnProAsnArgLeuLeuLeuLeuLeuGlnGlyPheLeuAspGlnAsnValHisPheAla	251
QY	2453	CACACAAACTCCGCGCTCCCAACCTATGCGAGACAGGAAACCTTAACCAAGCTCCAGATC	251
Db	252	HisThrSerIleLeuLeuSerPheLeuValArgAlaIleIleLysProTyrAspLeuGlnIle	271
QY	2513	TACCCCAACGAGACACAGATATTCGCTGGCCGAGTCCGCGACAGACACTGACAGACG	257
Db	272	TyrProGlnGlnArgHisSerIleArgValProIleSerGlyGlnHisTyrGlnLeuHis	291
QY	2573	TTGCTGCACTTTCTACAGGAATACCTC	2599
Db	292	LeuLeuHisTyrLeuGlnGlnIleAsnLeu	300

Db 387 GluTyrIleThrArgAlaGlyPhePheSerAspGlyThrThrValThrValGlnValMet 406
 Oy 1118 GAGCGGGCCAGCAGTGGCTCCAGTCCCTCCCTCCCGGGCCCGCTTCATCCCGAGC 1177
 Db 407 SerArgAspGlnAlaGlnCysSerLeuLeuLeuIleProTyrThrAspPheLeuPro 426
 Oy 1178 ACAGAG----- 1183
 Db 427 GluGluLeuGlyGlySerIleLeuGlnAspAsnLeuGlnLeuSerThrAspLeuAsnMet 446
 Oy 1184 -----AATGAGAGCAGCAGCGCTACCGCTCCAGAGCTGCCAG----- 1225
 Db 447 GlyValThrAspAspIysSerHisGluGluIleThrMetGluIysProProArgGlyLysLeu 466
 Oy 1226 -----AATGCCAGCGCTATGGTGTAGCAGAGCGTACAGACCGCTGTGATCAATGTT 1279
 Db 467 ArgGlyThrValGln-----IleHisLysAlaArgAsnAspTyrTrpIleAsnThr 483
 Oy 1280 CATGACATCTTCTATCCCTTCCCATCAGAGGAGAGAGCAGAGCTGC---TTTCTC 1336
 Db 484 HisAsnAlaIleTyrProLeuLysIleThrAspGluGlnHisProMetCylGluPheIle 503
 Oy 1337 CGCGCCATGATGCAGACAGCGGCTTCCCATTTGTACAAGTCACCGCGTTTAA 1396
 Db 504 TyrCysLeuGluIysProAsnGlySerCys---LeuAlaLeuIleSerAlaGluLeuAsp 522
 Oy 1397 TCCGAGCGCTTCGATGGAGTGGAGTGCAGCCCTTCAGCCCGGGAAGATTTAAAGTCCCC 1456
 Db 523 GlnAsnGlyTyr-----CysArg 528
 Oy 1457 ATTAAAGAGATTTGCTCTGACAGCGGTGATGGAGAGTTTGGCAGCGAGCTCC 1516
 Db 529 HisThrGluGluLysLeuLeuMetAlaGlnAsnPheSerIleAsnLysSerMetCyl--- 547
 Oy 1517 AAGATCTGGTCAATGAGAGACCAAGCTGTGATCTTCAAGGCGCAGAGACAGCGCG 1576
 Db 548 ---IleValValAspGluValArgGluLeuValTyrTyrValAlaAsnGluSerHisPro 566
 Oy 1577 CTGAGACGACGCTCTACGCTGTGATGAGCGGCGCGCGGAGATGTCAGCTCAGC 1636
 Db 567 ThrIleTyrAsnIle---CysValSerHisTyrArgThrGlyGlnHisAlaGluLeuThr 585
 Oy 1637 ACGCGCGGCTTCCCATGCTGCTCCATGAGCGCAAC-----TTC 1678
 Db 586 GluSerGlyIle-----CysPheLysSerGluArgAlaAsnGlyLysLeuAlaLeu 602
 Oy 1679 GACATG-----TTGTCAGCCACTACAGCAGCGTACAGCGCGCTGCGTGCAC 1729
 Db 603 AspLeuAspHisGlyPheAlaCysTyrMetThrSerValGlySerProAlaGluCysArg 622
 Oy 1730 GTTACAAGCTGAGCGCGCGCGAGAGACCCCTGCAAGACAGCGCGCTTGGGCT 1789
 Db 623 PheTyrSerPheArgTyrPylsGlnAsnGluValLeu-----ProSerThrValTyrAla 640
 Oy 1790 AGCATGTAGAGAGCGAGCTGCGC-----CCGAT-----TATGTTCCCTCAGAG 1837
 Db 641 AlaAsnIleThrValSerGlyHisProGlyGlnProAspLeuHisPheAspSerProGlu 660
 Oy 1838 ATCTTCATTTCCACAGC---CGCTCGATGTGGCGCTACCGCATGATCTACAAGCC 1894
 Db 661 MetIleGluPheGlnSerLysThrGlyLeuMetHisTyrAlaMetIleLeuArgPro 680
 Oy 1895 CAGCGCTTGCAGCGGAGAAAGACAGCCCGCTCTTGTATGAGAGCGCGCCAG 1954
 Db 681 SerAsnPheAspProTyrLysLysTyrProValPheHisTyrValTyrGlyLysProGly 700
 Oy 1955 GTGAGCTGTGATTAATCTCTCAAGCATCAAGTACTAGGCGCTCAACACACTGGCC 2014
 Db 701 IleGlnIleValHisAsnAspPheSerTrpIleGlnTyrIleArg-----PheCys 717
 Oy 2015 TCCCTGGGCTACGCGCTGTGTGATGAGCGAGCGGCTCTCAGCGAGGCGTTCGG 2074
 Db 718 ArgLeuGlyTyrValValValPheIleAspAsnArgLysSerAlaHisArgGlyIleGlu 737

Oy 2075 TTGCAAGGCGCCCTGAAAAACCAATGGGCGCAGTGGATCGAGACAGGTTGGAGGCG 2134
 Db 738 PheGluArgHisIleHisLysLysMetCylThrValGluValGluAspGlnValGluGly 757
 Oy 2135 CTGCAGTTCTGGGCGCAGAGATAT---GGCTTCATGACATGAGCGCGAGTTGCCATTCAT 2191
 Db 758 LeuGlnMetLeuAlaGluArgThrCylGlyPheMetAspMetSerArgValValHis 777
 Oy 2192 GCGTGGCTTACGGGGGCTTCCCTGCTGCTCATGGGCTTAATCCAAAGCGCGAGTGTTC 2251
 Db 778 GlyTyrSerTyrGlyGlyTyrMetAlaLeuGlnMetIleAlaLysHisPheAsnIleTyr 797
 Oy 2252 AAGTGGCCATCGGGGCGCGCGCTGACACCGCTGAGAGCGCTACAGACAGGTAACCT 2311
 Db 798 ArgAlaAlaIleAlaGlyGlyAlaValSerAspTrpArgLeuTyrAspThrAlaTyrThr 817
 Oy 2312 GAGCGCTACATGAGAGCTGCTCGAGAAACAGCAGCGCTATGAGCGGCTTCCGTGGCC 2371
 Db 818 GluArgTyrMetGlyTyr-Pro---LeuGluGlnHisValTyrGlyAlaSerSerIleThr 836
 Oy 2372 CTGCAGTGGAGAGCTGCCCAATGAGCGCGCTTGTATCTTCACGCGCTTCTG 2431
 Db 837 GlyLeuValGluLysLeuProAspGluProAsnArgLeuMetLeuValHisGlyLeuMet 856
 Oy 2432 GAGCAAAACGTGCACTTTTTCACACAACCTTCTGCTGCCAATGATCGAGACAGG 2491
 Db 857 AspGlnAsnValHisPheAlaHisLeuThrHisLeuValAspGluCysIleLysLysGly 876
 Oy 2492 AAACCTTACAGCTCCAGATCTACCCCAAGAGAGACACAGATTCGCTGCCCGCGAGTGC 2551
 Db 877 LysTrpHisGluLeuValIlePheProAsnGluArgHisGlyValArgAsnAsnAspAla 896
 Oy 2552 GCGCAGCACTAAGATCAGCTGCTGCACTTTCTACAGGA 2593
 Db 897 SerIleTyrLeuAspAlaArgMetMetTyrPheAlaGlnGln 910

Search completed: December 12, 2002, 11:51:46
 Job time : 108 secs

Db 379 ProHiscysValSerLeuTyrLysLeuSerProGluAspProthrcysLysThr 398
 QY 1775 CCCCCTTCGGGTGACATGATGAGGAGGAGCCAGCTGCCGCCCGGATATATCTCTCA 1834
 Db 399 LysGluThrPalatThrLileuAspSerAlaGlyProLeuProAspTyrThrPro 418
 QY 1835 GAGATCTTCATTTCCACAGCCCGCTGATGTCGGCTTACGAGCATGATACAGGCC 1894
 Db 419 GluIlePheSerPheGluSerThrThrGlyPheThrLeuTyrGlyMetLeuTyrLysPro 438
 QY 1895 CAGCCCTTGACAGCCAGGAGAGACACCCAGCTCTCTTGTATATGAGGCCCGCAG 1954
 Db 439 HisAspLeuInProGlyLysTyrLysTyrProThrValLeuPheIleTyrGlyLysProGln 458
 QY 1955 GTGACGTGTGATATACCTCTCAAGGATCACTGAGTGGCTTACAGCACTGAGCC 2014
 Db 459 ValGlnLeuValAsnAsnArgPheLysGlyValLysTyrPheArgLeuAsnThrLeuAla 478
 QY 2015 TCCCTGGGTACGCGCGGTGTGTGATTGACGCGAGGGCTCTGTACAGGAGGCTTCGG 2074
 Db 479 SerLeuGlyTyrValValValIleAspAsnArgLysSerCysHisArgLysLeuLys 498
 QY 2075 TTGAGAGGGCCCGAAACCAATGGCGGACAGTGGAGATGAGAGACAGGTGGAGGGC 2134
 Db 499 PheGlnGlyAlaPheLysTyrLysMet----- 507
 QY 2135 CTGAGTTCGTGGCGGAGATGATGCTTCATGACGTGAGCCGAGTTCATCCATGAGC 2194
 Db 507 ----- 507
 QY 2195 TGGTCTACAGGGGCTCTCTCTGCTCAGTGGGCTATCCACAGCCCGAGTGTTCAG 2254
 Db 507 ----- 507
 QY 2255 GTGCGCATCGGGGTGCGCCGCTGACGCTGTGATGAGCTTACGACAGAGTACAGTGA 2314
 Db 508 ValAlaIleAlaGlyAlaProValThrLeuThrIlePheTyrAspThrLysTyrThrGln 527
 QY 2315 CGGTACATGAGACGCTCTGAGAACACAGCAGCTATGAGGGGCTTCCTGAGCCCTG 2374
 Db 528 ArgTyrMetGlnLysProAspGlnAsnGlnGlnGlyTyrTyrLeuLysValAlaMet 547
 QY 2375 CACGTGAGAGAGCTGCCAATGAGCCCAACGCTGTGATGCTTATCCGACGCTTCCTGGAC 2434
 Db 548 GlnAlaGluLysPheProSerGluProAsnArgLeuLeuLeuLysLysLeuAsp 567
 QY 2435 GAAAGCTGACATTTTCCACACAACTCTGCTCCCACTGATCCGAGCAGAGGAAA 2494
 Db 568 GluAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeuValArgAlaGlyLys 587
 QY 2495 CCTTACAGCTCCAGATCTACCCCAAGAGACAGATGTCGCTGCCCGAGTCCGGGC 2554
 Db 588 ProTyrAspLeuGlnLysTyrProGlnGluArgHisSerIleTyrValProGlnSerGly 607
 QY 2555 GAGCACTATGAGTCACTGCTGCACTTTCTACAGAAATACCTC 2599
 Db 608 GlnHisTyrGlnLeuHisLeuHisTyrLeuGlnGlnAlaSerLeu 622

RESULT 7
 075868
 ID 075868 PRELIMINARY; PRT: 432 AA.
 AC 075868;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE R33083_1 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,

RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Dangnan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
 RA Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
 RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 RA Arellano A., Saunders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 RA Olsen A.S., Garrino A.V.,
 RT "Sequence analysis of a 2.5 kb region in 19p13.3."
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005783; AAC62840.1;
 FT NON_TER 432
 SO SEQUENCE 432 AA; 48595 MW; 64E2B85BE0523A7E CRC64;

Alignment Scores:
 Pred. No.: 1,07e-115 Length: 432
 Score: 3859.50 Matches: 354
 Percent Similarity: 97.52% Conservative: 0
 Best Local Similarity: 97.52% Mismatch: 0
 Query Match: 38.78% Indels: 9
 DB: 4 Gaps: 2

US-09-976-674-4 (1-2617) x 075868 (1-432)

QY 11 ATGCCACACCGGGACCCCAAGCCGACGAGCGAGCGACGCCGACAGATGACCG 70
 Db 72 MetAlaThrThrGlyThrProThrAlaAspArgGlyAspAlaAlaIleThrAspPro 91
 QY 71 GCGCGCGCTCCAGGTGAGAGACACGCTGGGAGCGGCTCCGAGCATCATACAGGC 130
 Db 92 AlaAlaArgPheGlnValAlaGlnLysHisSerThrPaspGlyLeuArgSerIleLeuHisGly 111
 QY 131 AGCCGCAAGTACTCGGGCTCATGTGTCAACAGGCGCCCACTTCCATTTGTGTCAG 190
 Db 112 SerArgLysTyrSerGlyLeuLeuValAsnLysAlaProHisAspPheGlnValGln 131
 QY 191 AAGAGGATGAGTCTGGGCGCCCACTCCACGCGCTCTACTACCTGGAGATGCCATGAGC 250
 Db 132 LysThrAspGlnSerLysProHisSerHisArgLeuTyrTyrLeuGlyMetProTyrGly 151
 QY 251 AGCCGAGAGAACCTCCCTCTACTCTGATGATTCGCAAGAGTCCGGAAGAGAGCTCTG 310
 Db 152 SerArgGlnAsnSerLeuLeuTyrSerGlnLeuProLysValArgLysGlnAlaLeu 171
 QY 311 CTGCTCTCTCTGGAAGACATGCTGATCATTTCCAGCCGACGCCACATGGGCTC 370
 Db 172 LeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaThrProHisHisGlyVal 191
 QY 371 TACTCTCGGAGAGAGAGCTGCTGAGAGGAGCGGAACGCTGGGGCTTCGGCATGCC 430
 Db 192 TyrSerArgGlnGlnGlnLeuLeuArgGlnArgGlyArgLeuGlyValPheGlyIleThr 211
 QY 431 TCTACGACTTCCACAGCAGAGATGGCTTCTCTTCCAGGCCAGCAGACAGCTCTTC 490
 Db 212 SerTyrAspPheHisSerLysSerLysLeuPheLeuPheGlnAlaSerAsnSerLeuPhe 231
 QY 491 CATGTGCGGAGCGGGCGGAAGACGCTTCAG-----GTGCC 529
 Db 232 HisCysArgAspLysGlyLysAsnGlyPheMetValSerProGlyProGlyCysValSer 251
 QY 530 CCTATGAACCGGTGGAATCAAGACCCAGTCTCAGGGGCCCGGATGAGCCCAATATC 589
 Db 252 PrometLysProLeuGlnLysThrGlnCysSerGlyProArgMetAspProLysIle 271
 QY 590 TGGCTTCGCGAGCCCTCTCTTCTTCTTCAATCAATACAGCAGCTGTGGTGGCCAC 649
 Db 272 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrValAlaAsn 291
 QY 650 ATGACAGACAGGAGAGAGCGGCGCTGACCTTCCACAGGATTATCAATGCTGCT 709
 Db 292 IleGlnThrGlyGlnGlnArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 311
 QY 710 GATGACCCCAAGTCTGGGGGTGGCCACTTGTCAATACAGAGAAAGTTCGACCGCTTC 769


```

Db 653 TygLylylproGlnAlaGlnLeuValAsnAsnArgPhyLyGlyAllyllystyrGheArg 672
OY 2000 CTCACACACTGGCCCTCCCTGGGCTACCGCCCTGGTGTGATTTAGCGGACGGGCTCTGT 2059
Db 673 LeuAsnThrLeuAlaSerLeuGlyTyValValAlaValAlaLeuAsnArgGlySerCys 692
OY 2060 CAGGAGGCTGGCTTCGAGGAGGCGCCGAAACAAATGGGACAGTGTGAGATTCAG 2119
Db 693 HisArgGlyLeuLeuysPheGlnGlyAlaPheLyTyLysMetGlyGlnLeuLeuLeuAsp 712
OY 2120 GACACAGTGGAGGCGCTGCATTCGTGGCCGAGAACTATGCTCATGACCTGACCCGCA 2179
Db 713 AspIleValGlnGlyLeuGlnIleuAlaSerArgTyAspPheLeuLeuLeuLeuLeuLeu 732
OY 2180 GTTCCCATTCATGGCTGGCTGTACGGGGCTTCCTCTGCTCATAGGGGCTAATCCACAAG 2239
Db 733 ValGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 752
OY 2240 CCCCAGCTGTTCAGAGTGCATTCGCGGCTGCGCCGCGCTGACGCTGTGATGGCTACGAC 2299
Db 753 SerAspIlePheArgValAlaAlaIleAlaGlyAlaProValThrLeuThrIlePheLeu 772
OY 2300 ACAGGAGTACATAAGGCTACATGAGACGCTCCCTGAGACACACACGCGCTATGAGGCG 2359
Db 773 ThrGlyTyThrGlnArgTyGlyMetGlyLysPheAspGlnAsnGlnGlnGlyTyTyLeu 792
OY 2360 GGTCCGCTGGCCCTGCACGCTGAGAGAGTGCACCATGACGACCCACGCGCTGTATCTCTC 2419
Db 793 GlySerValAlaMetGlnAlaGlnLysPheProSerGlnProAsnLeuLeuLeuLeuLeu 812
OY 2420 CAGGCTCTCTGAGCAAAACGTGCATTTTTCACAAATCTCTCTCTCTCTCTCTCTCTCT 2479
Db 813 HisGlyPheLeuAspGlnAsnValHisPheAlaHisThrSerIleLeuLeuSerPheLeu 832
OY 2480 ATCCGACACAGGAACTTCACAGCTCCAGATCCAGATCCCAAGACAGACAGATTTGGC 2539
Db 833 ValArgAlaLeuLysProLysArgLeuGlnIleTyProGlnGlyArgHisSerIleArg 852
OY 2540 TGCCCGAGTGGCGGAGACATGAGTCAAGTCAAGTCTCTCTCTCTCTCTCTCTCTCTCT 2599
Db 853 ValProGlnSerGlyGlnHisTyGlnLeuHisLeuLeuHisTyLeuGlnGlnLysLeu 872

RESULT 4
OY 075273 PRELIMINARY: PRT: 508 AA.
AC 075273;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN 11]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schulze K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Altix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Komyashki B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.,
*Sequence analysis of a 2.5 kb region in 19p13.3.
RL SMDIMTED (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005594; AAC33801.1;
DR MEROPS: S09.019;
DR InterPro: IPR002410; Pro-annoptase.
DR InterPro: IPR000379; Ser-estrs_site.
DR PRINTS: PRO0793; PROAMNOPTASE.
FT NON_TER 1 1

```

```

SO SEQUENCE 508 AA: 57750 MW: 2F4D6645BE2D2C89 CRC64;
Alignment Scores:
Pred. No.: 3,36e-163 Length: 508
Score: 2572.00 Matches: 493
Percent Similarity: 91.34% Conservative: 3
Best Local Similarity: 90.79% Mismatches: 3
Query Match: 53.64% Indels: 44
DB: 4 Gaps: 5

US-09-976-674-4 (1-2617) x 075273 (1-508)
OY 998 ATGCTGTGACCCAGGAGAGAGAGTGTGTCAGCCCTTCAGCTGCTCCCGAAGCTG 1057
Db 1 ILeuValSerThrGlnGlnLysGlnLeuValGlnProPheSerLeuProLysVal 20
OY 1058 GAGTACATGCCAGGCGCGGCGGACCCGAGTGCAGAAATACGCTGGCCCATGTCCTG 1117
Db 21 GlnTyThrIleAlaArgAlaGly-----AlaThrPalaMetPheLeu 33
OY 1118 GACGCGCCACACAGTGGCTGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1177
Db 34 AspArgProGlnGlnIleThrLeuGlnLeuValIleLeuLeuProPheAlaLeuPheIleProSer 53
OY 1178 ACAGAAATGAGAGACAGCGGCTACGCTCTGCSAGAGCTGTCCCGCAAGTGTCCAGCGC 1237
Db 54 ThrGlnAsnGlnGlnGlnArgLeuAlaSerAlaArgAlaValProAlaGlnValGlnPro 73
OY 1238 TATGTGCTTACAGAGAGTCAACAAGCTGTGATCAATGTTCATGACATTTCTATCC 1297
Db 74 TyValValTyGlnGlnIleValThrAsnValThrIleAsnValHisAspIlePheTyPro 93
OY 1298 TTCCCCCAATGAGAGGAGAGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1357
Db 94 PheProGlnIleSerGlnGlyGlnAspLeuLeuLeuLeuLeuLeuAlaAsnGlnLysTyThr 113
OY 1358 GCGTCTGCAATTTGTACAAATGCAAGTCAAGCTTTTAAATCCAGGCTACATTTGAGAT 1417
Db 114 GlyPheCysHisLeuTyThrGlyValThrAlaValLeuLysSerGlnGlyTyAspTrpSer 133
OY 1418 GACCCCTTACGCGCGGAGAGATGAATTTAAGTCCCATTAAGAAAGATTCGCTG 1477
Db 134 GlnProPheSerProGlyGlyGly-----GlnGlnSerLeu 145
OY 1478 ACCAGCGTGAATGGAGAGTGTGGCGAGCGAGCTCAAGATCTGGTCAATGAGAG 1537
Db 146 ThrAsnAla-----IleThrValAsnGlnGln 154
OY 1538 ACCAAGCTGTGTAATTCAGAGGACCAAGACAGCGCGCTGAGACACACATCTACGTG 1597
Db 155 ThrLysLeuValTyThrGlnIleTyThrLysAspThrProLeuGlnHisHisLeuTyVal 174
OY 1598 GTACAGTATGAGCGCGCGCGAGATCTACGCGCTACACAGCGCGGCTCTCCATAGC 1657
Db 175 ValSerTyGlnAlaAlaGlyGlyIleValAlaArgLeuThrThrProGlyPheSerHisSer 194
OY 1658 TGCTCATGACCCAGAACTTCGACATGTTCTGCAGCCATACACACAGCTGAGACGCG 1717
Db 195 CysSerMetSerGlnAsnPheAspMetPheValSerHisTySerValSerThrPro 214
OY 1718 CCTGTGCTCAGCTGTACAGTGTAGCGGCGCGCGAGCGAGCGAGCGCGCTGCAAGACAGCC 1777
Db 215 ProGlyValHisValTyThrLysLeuSerGlyProAspAspAspProLeuHisLysGlnPro 234
OY 1778 CGCTTGTGGCTAAGATGAGGAGCGACAGCTGCGCCCGGATTAATGTTCTCCAGAG 1837
Db 235 ArgPheThrAlaSerMetGlnAlaAla-----Lys 245
OY 1838 ATCTCATTTTCACACAGCGCGCTGAGTGTGCGGCTACAGGCAATGATCAAGCCCGAC 1897
Db 246 IlePheHisPheHisThrArgSerAspAlaArgLeuTyGlnMetIleTyLysProHis 265
OY 1898 GCCTTGACAGCGAGGAGAGACACCCACAGCTCTTTTGTATATGAGGAGCCCGACGTC 1957

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:03:27 : Search time 23 Seconds
(without alignments)
3607.131 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATTTGTTATRGDAATDDP.....CPESGEHYEVTLHFLQEYL 863

Scoring table:

BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	934.5	20.1	931	2	TJ2919
2	805.5	17.3	738	2	A87516
3	774	16.7	741	2	JC5142
4	585.5	12.6	711	2	S66261
5	508.5	10.9	793	2	T41703
6	487.5	10.5	766	1	CDHU26
7	484.5	10.4	792	1	A59914
8	480	10.3	760	1	S23752
9	477	10.3	779	2	T25173
10	467	10.1	779	2	T25173
11	445.5	9.6	803	2	I68600
12	445.5	9.6	865	2	I54331
13	442.5	9.5	803	2	A41793
14	432.5	9.3	818	1	A30107
15	417	9.0	931	2	A49737
16	394.5	8.5	795	2	F82858
17	333.5	7.2	829	2	T19514
18	313.5	6.7	759	2	I38593
19	286.5	6.2	683	2	E87495
20	270.5	5.8	709	2	B82580
21	262	5.6	743	2	T37700
22	249.5	5.4	657	2	E70025
23	240.5	5.2	642	2	C71137
24	234	5.0	631	2	H75007
25	215.5	4.6	732	1	UC4655
26	212.5	4.6	732	1	S07624
27	204	4.4	569	2	S74053
28	203	4.4	622	2	F71174
29	202	4.3	591	2	H72474

30	194.5	4.2	732	1	JU0132	acylaminoacyl-pept
31	194	4.2	536	2	F90299	acylaminoacyl-pept
32	191.5	4.1	582	2	D72636	probable acylamino
33	191	4.1	606	2	T35378	probable acylamino
34	191	4.1	761	2	S44807	probable peptide h
35	189	4.1	632	2	E75057	F44B9.1 protein -
36	187.5	4.0	629	2	T15945	peptidase PAB1418
37	186.5	4.0	667	2	A87711	hypothetical prote
38	185	4.0	674	2	B84381	prolyl oligopeptid
39	180.5	3.9	658	2	F83397	acylaminoacyl-pept
40	178.5	3.8	605	2	T35975	probable peptidase
41	176.5	3.8	819	2	B87580	probable acyl-pept
42	174.5	3.8	572	2	F72455	conserved hypothet
43	173.5	3.7	955	2	F84914	probable acylamino
44	167	3.6	265	2	B84063	hypothetical prote
45	165.5	3.6	705	2	JX0194	hypothetical prote
						prolyl oligopeptid

ALIGNMENTS

RESULT 1									
TJ2919									
hypothetical protein K02F2.1 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999									
C:Accession: TJ2919									
R:Maggi, L.; Goela, D.									
A:Description: The sequence of C. elegans cosmid K02F2.									
A:Reference number: Z21246									
A:Accession: TJ2919									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-931 <MAG>									
A:Cross-references: EMBL:AF043699; PIDN:AAB97564.1; GSPDB:GN00019; CESP:K02F2.1									
A:Experimental source: strain Bristol N2; clone K02P2									
C:Genetics:									
A:Gene: CESP:K02F2.1									
A:Map position: 1									
A:Introns: 58/3; 82/2; 131/2; 178/2; 275/3; 322/2; 404/2; 441/1; 464/2; 486/3; 528/2;									
Query Match									
Best Local Similarity 29.5%; Pred. No. 1e-63; Length 931;									
Matches 264; Conservative 142; Mismatches 322; Indels 167; Gaps 33;									
QY	72	IYLLGMPYGSRENSLLYESEIPKVV-----RKPALLLSWKOMLDHFOATPHHGVS	SR	123					
DB	82	KVAISSVGTNTQSIIFSVTIPELYEKAQVADRKTELKSGYVNDSTYRMSCRTTPSA		141					
QY	124	EEELREKRRKLGVF-GITSYDFHSESGLELFGQASNSLFH-----		161					
DB	142	EFTLQECRQRSGVYVIGISDYEL--RNGKMILMAGDQLFRYNPLNLAALPIAVDDQSS		199					
QY	162	-----CRDGGKNGFVSPMKPLEIK--TQCSGRPRDP-----		192					
DB	200	TEPMDISGSGITSQSGKCSNEAPQSSVPPVRIPIKPTTSTKPPATAPANNFVSSAK		259					
QY	193	ICPADPAFFSFTNNSDLWVANITETGEERLFTCHQGLSNVLDPPKSAGVATFVIOEEDR		252					
DB	260	VCPADSSLLAYLVNKKQVYI-----EKNGKTIHRTSSN--SKHITNGVPSYIVDELER		310					
QY	233	FTGYWMCPTASWEGSGELTLRLIYEEVDESEVEV---IHVSPALREKRTDSTYRPT		308					
DB	311	FEGIWV-----SES-KT-RLIYEHVNEEKVAESQFCVNDPPVA-----PKKYPRA		354					
QY	309	GSKNKIKMLKLAFFQTDQSGKTVSQEKLQVPSFLPKVYIARAGTRDGKXAMNF		368					
DB	335	GKKNAVSTLRWVILE--NGKAYDVLKDEV--IYKHCFEYITIRAGFSGGTTWQOV		409					
QY	369	LDPPQWOLVLLPPLAFIPSTE-----NEEOLASARAVPR--		405					
DB	410	MSRDQAQCSLLIPLYTDFLLPELGSTIKEDNLQJLSTDLNMGMVMDKSHETMKEKPRGX		469					

```

QY 406 ---NWOPRVVYEEVYNWNIWNIHVDIYFPEPPOSQGEDEL-C-FLEANCKTGFCYLIVYTAU 461
D 470 LRGIVQ---IHKARNDVYNINHNATYPLKIDDEHPMEFFIYCLEKPNQSC-LALISAEU 525
QY 462 KSGQIDWSEPPSPGDEFEKCPRIKEIALTSGEWEVLARHGSKIWNETKLVYFOGTYDT 521
D 526 DONGY-----CRHTEKLLMAENFNSINKSMG--IYDEVEKLELYYYANESH 569
QY 522 PLEHNLVYVSEAGELVRLTPGSHSCMSQN-----FDM---FVSHSVSTPECV 572
D 570 PTEWNT-CVSHYRFGQHAQLTESGI--CFKSERANGKLADLDHGFACMYMVSQSPABC 625
QY 573 HUYKLSGPDDBPLHKOPRFMSMMEASCP--PD--YVPPETIHNHT--RSDVRLKMYK 627
D 626 RFYSRWKENEVL--PSRYVAANITVSGHPQOPDLHFDSPKEMIERQSKTGIMHTAMILR 683
QY 628 PHALOPGKKHPTVLFVYGGPOVQVLYNNSFGKIKYLRILTASLGAAYVYIDGRSGQRCU 687
D 684 PSNFDYPRKKYVFHFHYVGGPGIQIWHNDEFSWIOYIR--FCRLGYVVFIDNRGSAHBGI 740
QY 688 RFEGALKQMGQVELEDOVEGLOFAVEKY-GFIDISRAVHGMSVGGFSLMLGLHNKPOV 746
D 741 EFERHIIHKMGITVEEDVQVEGLQMLAERTGTFMDKSRVYVHWSGCTMALDQMIKHNPI 800
QY 747 FKVAIAGAPVTWMAVYDTGYTERYMDVPENNOHGYEAGSVLHVEKLPNEPNRLLILHGF 806
D 801 YRAAIAGAVSDWRLYDTRATERYWGYF-LEEHVYGASSITGLVEKRLDDEPRRLMTVHGL 859
QY 807 LDENVNHFHTNPLVSQLFRACKPQOLOIYPNRHSIRPDESEHNEVLLNHLLOE 861
D 860 MDENVNHFHLTHLDECIKKGKWHMELVIFPNRHRGVRNNDSIYDARMKTFPAQO 914

```

```

RESULT 2
A:7516
dipeptidyl peptidase IV [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87516
R:NIEMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HELDBERG, J.
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFL, D.H.; KOLON
N, J.; ERMOLEVA, M.; WHITE, O.; SELZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4133-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249, MUID:21173698, PMID:11259647
A:Accession: A87516
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-738 <STO>
A:Cross-references: GB:AE005673; NID:G13423647; PIDN:AAK24125.1; GSPDB:GN00148
C:Genetics:
;:Gene: CC2154

```

	Query Match	17.3%	Score 805.5;	DB 2;	Length 738;
	Best Local Similarity	28.8%;	Pred. No. 6.8e-54;		
	Matches 227; Conservative 118; Mismatches 308; Indels 135; Gaps 25.				
Oy	67 PHSRLTYL-GMPGSRNSLLYSEI----KYAKKELLLSMKOMLDHFQATPHHGY 121	: : : : :	:	:	:
Dd	47 PDGRVVTYLKKRPEANITODLMADVKGEPRYRLDLSALSGDEKL----- 93	: : : :	:	:	:
Oy	122 SREELLERERRRGVFITYSYDFHESGLFLFQASNSLFHCRODGKNCFMVSPPMKPLEIK 181	: : : :	:	:	:
Dd	94 SEARKARRERARVRSAKGIVETSMDRQGRIPLPDGDLY--LDNAVAG-----KITRL 144	: : : :	:	:	:
Oy	182 TQCGSPMDPKICPADPAFFSFINNSDLAVANIETGEERRLTFCHQGLSNVLDPKSAGV 241	: : : : : :	:	:	:
Dd	145 TETPGDEVDAVASPKG-CGVSVYVRDNLYIRPVAGGAFTALTIDDK-----DALSPGV 196	: : : : : :	:	:	:
Oy	242 AIFYIQEFDKFCTGCWPCPTASWBEGSLKTLRLILEVDSEVEVINYPSALEERTKD 301	: :	:	:	:
Dd	197 AEPIVGDELDTFTGTWMSPDSS-----RLVTVNDSGVDIVRADIGPGATVV 246	: :	:	:	:

[illegible]

```

RESULT 3
JC5142
X-Pro dipeptidyl1-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C:Date: 31-Jan-1997 #sequence,revision 27-Feb-1997 #text,change 20-Jun-2000
C:Accession: JC5142
R:Kabashima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl1 peptidase IV from Xanthomonas maltophilia: Sequencing and express
A:Reference number: JC5142; MUID:97164011; PMID:9010738
A:Accession: JC5142
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <KAB>
A:Cross-references: DDBJ:D83263; NID:g1753196; PIDD:BA11872.1; PID:g1753197
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl1
C:Superfamily: dipeptidyl1-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted <TM>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted

```

Query Match 16.7%; Score 774; DB 2; Length 741;
Best Local Similarity 29.3%; Pred. No. 1.9e-51;
Matches 226; Conservative 125; Mismatches 275; Indels 146; Gaps 25;

OY 108 MDLHQATPHHGIVSRREELLRRKRLGVF-GITISDYDHSSEGLFLFOASNS--LFHQRD 164
:::|||||:::|||||:
DB 79 LVDSKVVLPGTETLSDEKKARRORIAAMGIVDYQSPDAQRLLFLPLGCELYIYYDKO 138

```

QY 165 GKGNGVSPMKPLEIKTQCSGPRMDPKICPADAFSFINNSDLVANIETGEERLTF 224
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 EGK-----AAVRLT---THGEGFATDAKLSPEKG-GEVSFIRGNRLMWIDIASGRQMOLT- 188
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 CHQGLSNVLDPKSAGVATFYIOEEFDFRTGYWMCPTASWESSEGLKRLRIIYEVDSE 284
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 -ADGSTT-----GNGIAFEVADEMDRHTGTWMAPDSS-----ATAVARIDESP 232
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 VEVIHVSPALER-----KTD--SYRPRTGSKNPKIATLAEFQTDOSGKIYSTQEK 336
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 VPV-----OKREYVADRTDVEQRYPAAGANVOYKGVISPAEQAOQWIDLGKE 284
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 ELVOPFSSLPKVEYIARAGWTRDQKYAMAFELDRPQOOLVLLPPLFTPSTENEROR 396
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 QDI-----YLAIVNM-RDPQHLSFORQSRDQKKLDV-----EVT 318
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 397 LASARAVRNQPYVYVEVTVNWIINVHDIFYPPPOSGEDELCLFRANECKTGCHLYK 456
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 LASN-----QORVLAHETSPTWPLHN-----SLRFLLDGSILWMSSE-RTGFQHDYR 364
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 457 VTAVLKSQGYDMSEPSGDEBEFKPIKEETALTSGEV---EVLARHSGKIWNVEETKLY 513
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 365 I-----DSRG-----KAAALYHGMSVDLDA-----VDEKAGLA 394
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 514 YFOGTRDPLEHHLVYVEAAGEIVRLTTPGFSHSCMSQNFDMFVSHYSSVSTPCVH 573
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 YFRAGIESARESOIYAVLQ--CGQPORLSKAPGMHSAFARASAYVWSMSNNSIPPOLE 453
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 YKSLSGPD-----DDPLHKOPREWASMEASCPDDYVPEIFHFRSDVR--LY 622
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 LFRANGEXIATLVENDLADPKHRYARY---REAR-----PVEFGTLTAADGKTPLN 502
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 623 GMVYKPHALOPGKKHPTLVFYGGPOVOLYNNSEFKI--KYLRLNLTSLGTVVYVYIDGG 681
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 YSVIKPAFPDPKRIYAVVYVYGGPASTVDSWPGRGDHLFNQYLAQGYVFSGLDNG 562
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 682 SCQGRLEFEGALKNQMGVEIEDQVEGLQFYAEKGYFTDLSRVAIHGMSYSGGLSLMGLI 741
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 TPRGRDRFGALYKQGVVEVADQRGVAMLKQ--PWVDPARIGVOGMSNGGYMLMLLA 621
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 742 HKPOYFKVAGAPYVVMAYDTGTERMDVPENNQHGEGYAGSVALHVEKLPNEPNRL 801
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 KASDSYACGVAGAPYTDGLYDSHYTERYMDLPARNDAGYREARVLTHEGLRSP---LL 678
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 802 ILHGLDNVHFEHTNFTVLSQILIRAGKPYOQIYPERHSIRCPSEGEHYEV 853
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 679 LIHGADNVLFTNSTSLMSALQKRGQPELMTYTPGAKHGUSGADALHRYRV 730
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4

```

S66261
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
C:Species: Flavobacterium meningosepticum
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66261
R:Kabatima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
Arch. Biochem. Biophys. 320, 123-128, 1995
A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene from F1
A:Reference number: S66261; MUID:95314307; PMID:7793970
A:Accession: S66261
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-711 <KAB>
A:Cross-references: EMBL:D42121, NID:9577283; PIDN:BA07702.1; PID:9577284
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase

```

```

Query Match 12.6%; Score 585.5; DB 2; Length 711;
Best Local Similarity 25.2%; Pred. No. 6.9e-37;
Matches 184; Conservative 130; Mismatches 266; Indels 147; Gaps 24;
QY 142 YDF-HSEGLFLFOASNSLFLHCRDGGKNGFVWSPMKPLEIKTQCSGPR---MDPKICPA 196
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 81 YTFNSDESKILLQKSSQSIYRHSFLGK-----FEVKKDLKSRTVSLNNANWIOEPRFSP- 134
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 DPAPFSFINNSDLVANIETGERLTFCHQGLSNVLDPKSAGVATFYIOEEPRFNGY 256
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 DQSKVAFIADNNLFYQDNLTKITQIT--TDGKKNELIN---GLGDWYEEERGHADY 188
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 WMCPTASWESSEGLKRLRIIYEVDSEVEVYHNV-----SPALERKTSYRPRGTS 310
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 189 QMKNAGD-----ALVFRPDERRYPEINITYQNLYP-----KIMTYKKPKAGE 233
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 311 KNPRTALKLAEFQTDOSGKIYSTQEKELVOPSSLPKVEYIARAGWTRDQKYAMAFEL 370
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 ENSAVTAVL--YQLSS-----GRSAOLNFGS 257
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 RPOQLOLVLLPPLFTPSTENERORLASARAVRNQPYVY-----YEEVTWV 420
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 SEKYIT-----POLF--OTNANDELIVANANRHNQKVDLLKNTKTAAVSKLFTETDNW 310
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 421 INVHDIYFPPOSGEDELCLFRANECKTGCHLYKVTAVLKSQGYDMSEPSGDEBEK 480
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 IETDNLTFEF-----LDDNSFLMASE-RDGHRLIYWDYAGK----- 346
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 481 CPKEEIALTSGEWELARHSGKIWNVEETKLYVFOGTRDPLEHHLVYVEAAGEIVR 540
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 --LKKQV--SKGDWEITINYG---YNPKTKEVYIQTTEKGSINKVYSKLININ-TGKTOL 397
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 541 LTPGFSHSCMSQNFDMFVSHYSSVSTPCVHYVTKLSGD-----DDPLHKOPRWA 593
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 LSNAGGNNSAARSKTFNFINTSTPAKYPTKYLKDANCKDKRELQNDNDLLNKK- 453
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 594 SMWEASCPDDYVPEIFHFRSDVRLYGMVYKPHALOPGKKHPTLVFYGGPOVOLYN 653
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 -----SDNIAEFITIPMAADQMANNAIKRKNPDPAKRYVPMFOYSGPSQOYA 505
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 654 NSFQIKYLRILTASLGYAVVYIDGRSCQGRLEFEGALKNQMGVEIEDQVEGLQFYA 713
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 NSMDGNGCIGWFMLMLQKQGLVYCVDRGRTGFGRTKYKTKYKTLKYEIEDQITAAKWL 565
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 714 EKYGFIDLSRVAIHGMSYSGGLSLMGLIHKPOYFKVAGAPYVVMAYDTGTERYMDV 773
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 NO-STYVDSRIGIFGWSGGYMASLMTKGADEVFKMGLAVATVWRFDSLYTERFLQ 624
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 774 PENNOHGEAGSVALHVEKLPNEPNRLILHGLFDENVFEHTNFTVLSQILIRAGKPYOQ 833
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 PQENNDQDGLNLSPTTYAKL---KGFLLIHGTADNHNHFQNSMESEALLQNKQFDEM 681
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 834 IYPERHST 842
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 682 AYPDKNHSI 690
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5

```

T41703
dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41703
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z22011
A:Accession: T41703
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-793 <MUR>
A:Cross-references: EMBL:AL031180; PIDN:CAA20138.1; GSPDB:GN00066
A:Experimental source: strain 972h-; cosmid c2E11 -chimeric
C:Genetics:
A:Gene: SPAC2E11.08
A:Map position: 1
C:Superfamily: dipeptidyl-peptidase IV

```

```

Query Match 10.9%; Score 508.5; DB 2; Length 793;
Best Local Similarity 25.4%; Pred. NO. 7.5e-31;

```


QY 272 TLRLIVEDESEVEVH---VSPALERKTDSDRYPRGSKNPRIALKLAEPOTDSOG 328
 Db 223 ---LAIQAFNDTEVPLEIETSYSDPSLQYPTVAVPIPKAGAVNTV--KFFVNTDLS 277
 QY 329 KIVSTOEKELVQPFSSLEPKVEYIARAGWTRDGKYAMAMFLDRPOWLQTLPLPALFIP 388
 Db 278 SVTNATSIQITAPASMLIGD-HYLCDTWTATQERIS-----LQWL----- 316
 QY 389 STENEORLASARAVPRNVQYV---YEYVNWVINVHDIFYPPOSEGDELCFLR 443
 Db 317 -----RRIONYSVMDCIDYDESSGRW----- 337
 QY 444 ANECKTGCHLYKTAVALKSGQYDMSEPEPSEDE-----FK 480
 Db 338 --NCLVARQHLEMTT-----GWGRFRSEPHFTILDGSEFKIISNEGYRHICYFO 388
 QY 481 CPIKEETALTSGEWEVLARHGSKIWNNEETKLYFOGKTDPLEHHLVVSYEAG--- 536
 Db 389 IDKDCFFETFGTWEVIGIEAL-----TSD-----YLYISNEYKGMFGG 428
 QY 537 -EIVRLTPPGFSHCSQNDMFVSHSYSTPCVAVYKL--SGP----- 580
 Db 429 RNLKYIQLIDTKYVTCLSCELNPERCOYYSFSEKRAYYOLRGSGPLPYTLHSSVND 488
 QY 581 -----DDPLHKQPRFMAEMEAASCPDYVPEIFHFHTRSDVRL-YGMIKPHALQ 632
 Db 489 KGLRVLEDNSALDK-----MLQWQ-----MPSKIDFTILLETWKWYMLPPH-FD 535
 QY 633 PGKHPYVLFYVGGPOVOLVNNSEFGKIKYRLN--TLASIGYAVV--IDRGSGCORGLR 688
 Db 536 KSKYIPLLDVYAPGSCQKADTVF-----RLNATYLASTENITIVASFDRSGSYQGD 589
 QY 689 FEGALKMOMQVEIEDOYEGE-QVVAEYGGTIDSRVAIHGWSIGGELISLGLHKPOVF 747
 Db 590 IMHAINRRLGTFEVEDDQIEAARQF--SKMGFVDMKRIAMQWSYGGVYTSVVLGSGSGVF 647
 QY 748 KVALAGAPVYVMAVYDGYTERYMDV--PENNGYEGASVALHVEKLPNPNRLILIHG 805
 Db 648 KCGIAPVAVSMWEYDYSTRYKMGCLPRPEPNLDHTRNISTVMSRAENF--KOVETYLTHG 705
 QY 806 FLDENVHEFTNPLVSOLIRAGKPYOLOIYENERSIRCPESGHEHYETLLHFOE 861
 Db 706 TADNVHFGQSAQISKALVDYGVDFQAMWYTDDEHGIASSTAHQIITYHMSHFITQ 761
 RESULT 7
 A39914
 dipeptidyl-peptidase IV (EC 3.4.14.5), membrane-bound form precursor - rat
 N:Alternate names: gp110; membrane glycoprotein 110K; OX-61
 N:Contains: dipeptidyl-peptidase IV, soluble form
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text, change 10-Sep-1999
 C:Accession: A39914; A33315; B33315; A60730; A42203; S38949; A31781
 R:Hong, W.; Doyle, D.
 Proc. Natl. Acad. Sci. U.S.A. 84, 7962-7966, 1987
 A:Title: cDNA cloning for a bile canalicular domain-specific membrane glycoprotein of rat
 A:Reference number: A39914; MUID:88068516; PMID:3479775
 A:Accession: A39914
 A:Molecule type: mRNA
 A:Residues: 1-792 <HON>
 A:Cross-references: GB:J02897; NID:g204463; PIDN:AAA41272.1; PID:g204464
 R:Ogata, S.; Misumi, Y.; Ikehara, Y.
 J. Biol. Chem. 264, 3596-3601, 1989
 A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA at
 A:Reference number: A33315; MUID:89123496; PMID:2563382
 A:Accession: A33315
 A:Molecule type: mRNA
 A:Residues: 1-37, 'A', 39-182, 'I', 184-331, 'T', 333-351, 'C', 353-393, 'V', 395-561, 'L', 563-623,
 A:Cross-references: GB:J04591; NID:g203973; PIDN:AAA41096.1; PID:g203974
 A:Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
 A:Accession: B33315
 A:Molecule type: protein

A:Residues: 1-20;35-54;427-443;505-509;511-520;530-538;593-600;602-608;618-627 <H02>
 R:McCaughan, G.W.; Mickson, J.E.; Crewick, P.F.; Correll, M.D.
 Hepatology 11, 534-544, 1990
 A:Title: Identification of the bile canalicular cell surface molecule gp110 as the ec
 quence.
 A:Reference number: A60730; MUID:90228896; PMID:1970322
 A:Accession: A60730
 A:Molecule type: protein
 A:Residues: 28-47, 'X', 50-53, 55-58 <MCC>
 R:Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
 Biochemistry 31, 2582-2587, 1992
 A:Title: Identification of the active site residues in dipeptidyl peptidase IV by aff
 A:Reference number: A42203; MUID:92190188; PMID:1347701
 A:Accession: A42203
 A:Molecule type: protein
 A:Residues: 'R', 625-630, 'X', 632-648 <OG2>
 R:Iwaki-Egawa, S.; Matanabe, Y.; Fujimoto, Y.
 Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
 A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl
 A:Reference number: S38949; MUID:94128239; PMID:7905271
 A:Accession: S38949
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 281-302 <IWA>
 R:Hong, W.; Doyle, D.
 J. Biol. Chem. 263, 16892-16898, 1988
 A:Title: Membrane orientation of rat gp110 as studied by in vitro translation.
 A:Reference number: A31781; MUID:89034185; PMID:3162821
 A:Accession: A31781
 A:Molecule type: mRNA
 A:Residues: 1-40 <H03>
 A:Comment: This protein is localized to the bile canalculus, which is the apical dom
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidyl-peptidase hydrolase; glycoprotein; homodimer; liver; serine prote
 F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM
 F:1-78/Domain: signal sequence #link MATS #status experimental <SIC>
 F:1-6/Domain: intracellular #status predicted <INT>
 F:7-28/Domain: transmembrane #status predicted <TM>
 F:29-79/Domain: extracellular #status predicted <EXT>
 F:29-34/Domain: propeptide #link MATS #status experimental <PRO>
 F:35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
 F:83;90;148;217;227;319;521;686/Binding site: carbohydrate (asn) (covalent) #status p
 F:631/Active site: Ser #status experimental
 F:709;741/Active site: Asp, His #status predicted

Query Match 10.4%; Score 484.5; DB 1; Length 792;
 Best Local Similarity 23.6%; Pred. No. 5.4e-29;
 Matches 205; Conservative 123; Mismatches 334; Indels 207; Gaps 36;

QY 63 DESGHSRLV---YLGPYSGRENSILY--SEIPKVKRKALLLLSKKOMLDRHQATPH 117
 Db 31 DEAAADSRRTYTLADYLNKTRFKVSYSLRWSDSEYLYKQENNTLL-----FNA--E 80
 QY 118 HGVSREBELLRERKRLVFCITSYDFH-SESGLELFOASNLFLPCRDGKNGFVSPMK 176
 Db 81 HG----NSITLENSTFEIIFGDSISDYSPDRFLVLEYNVKKMRISYASYSITDNL 136
 QY 177 PLEIKTQCSGPRMDPKICPADPAF--FSFINNSDLWVANIETGEERRLTFCGGLSNVID 235
 Db 137 KROLITEKRIPINNQIMWISQEGHKLAVYKWNNDIYKIEPHLPSHRT--STGKENVIFN 194
 QY 236 PKSGVAFVFIQEE-FDRFTGYWGCPTASWGSSEDLKRLITLIEVDSSEVEVH---VP 291
 Db 135 ----GINDVYEEETFGAISALWMSNCTF-----LAIQAFNDTEVPLEIETSYSDPS 240
 QY 292 SPALERKTDSDRYPRGSKNPRIALKLAEPOTDSOGKIVSTOEKELVQPFSSLEPKVEY 351
 Db 241 DESIQYPTWMLPYPKAGAVNTV--KFFVNTDLSSTTTTIPQITAP--ASVTTGGHY 297
 QY 332 IARAGWTRDGKYAMAMFLDRPOWLQTLVPLPALFIPSTENEORLASARAVPRNVQYV 411
 Db 298 LCDVAVWSEDRIS-----LQWL-----RRIONYS 321

Query Match	10.3%	Score 477	DB 2	Length 799
Best Local Similarity	24.2%	Pred. No. 2.1e+28		
Matches 217; Conservative	134	Mismatches 315	Indels 230	Gaps 47

```

Db      628  RSAIYGQLGTLVEEDQIKAIKVLYRLRIYHLLDARRRAVAGWSVGGMFLISMNVEAEQGF 687
Qy      748  KVALGAPRTVMMAADTGTERTM-DVPENNOHGCEAGSVAHLVKLPN-EPNRLLILHG 805
       1 : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      688  KCASVAPPTNAAYATATRYTERMGDAP-----LESYSDTKKLDLNFSTRLLIMHG 739
Qy      806  FLDEVNHFEHTMFIVLSQILRACKPKYQLOLYIPNERSIRSCPESEGEHEVILLHFLOE 861
       1 : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db      740  LDDNVNHFQNSALLIDELONRGVDPLMYYPINOAHSLSRST--HYVGKMTHFLKQ 793

RESULT 10
T25173
hypothetical protein T23F1.7a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T25173
submitted to the EMBL Data Library, October 1996
A:Reference number: 219990
A:Accession: T25173
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-779 <WIL>
A:Cross-references: EMBL:T281129; PIDN:CAB03411.1; GSPDB:GN00023; CESP:T23F1.7a
A:Experimental source: clone T23F1
C:Genetics:
A:Gene: CESP:T23F1.7a
A:Map position: 5
A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 683/1; 723/1
C:Superfamily: dipeptidyl-peptidase IV
```

[illegible]

Db 444 LADL-----SRNSAYCISCSIKNCTWAOQMDOMKTAIVSCGPAAPHTAIVLITRMDSK 500
QY 585 LHKQ-----PRFASMAEASCPDVPPELFHFHTSDVLYGMKYKPHALQGGKKH-- 637
Db 501 KTEHANLLYDTTYONRVEEGLP--VLIKETIKISDPDALIKLSTI--PKDIYVRDHQA 556
QY 638 -PVLFFYGGPOVOLVNNSEK-GIKYLRLLNTLASGYAVVY-IDGRSGCGRLFEQALK 694
Db 557 IPLVHYGGNDQNTKEAQIGIE----EYVASASQAAILRIDRGSGGGMKRYSAIY 612
QY 695 NQMGVEIEDQEGQFVAEKY-GFIDLSRYAHGWSYGFSLMGLIHKP-QVFXVAIA 752
Db 613 GQLCTVEVEDQIKAIKYVLRILYRHLIDARVAVFGWSYGFMTLSMVEAEQEFKCAVS 672
QY 753 GAPYVMAVDQTYTERYM-DVPENNQHGYSAGSVALHVEKLPN-EPNRLLILGHFDEN 810
Db 673 VAPYVMAVDQTYTERYM-DVPENNQHGYSAGSVALHVEKLPN-EPNRLLILGHFDEN 810
QY 811 VHFHTNELYSQLIRAGKPYOQIYPERHSIRCPESGEHYEVTLLHFLQ 861
Db 725 VHFQNSAILDELQNRGYDFLWYPMQAHSLSRST--HYVGKMTHFRLQ 773

RESULT 11
168600
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: I68600
R:Yokocani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-relat
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I68600
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-803 <RES>
A:Cross-references: GB:M96860; NID:g306707; PIDN:AAA35761.1; PID:g306708
C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.68; Score 445.5; DB 2; Length 803;
Best Local Similarity 23.88; Pred. No. 5,7e-26;
Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;
QY 150 LFLQASNSLFCHCDGKNGFMVSPMKPLEIKTQCSGRMDPKICPADPAFFSTINSD 209
Db 193 IFIRE--NNIYYCAHVQKQAIRV----- 213
QY 210 WVAIETGEERRLTFCHQGLSNVLDLDPKSAVATFVIOEEDRFTGYWMCPTASMEGSEG 269
Db 214 ----VSTGKEGVI---YNGLSDWLYE-----EILKTHIAHWSPDGT----- 249
QY 270 LKTRILYEEDSEVEVYIHPV-----PALERKTDSTYPTGSKNPYIAKLAIEFQ 323
Db 250 ----RLAYAAINDSRVPMELPTGTSTYPTVK-----PYHYPRAGSNPISLSHVI--- 297
QY 324 TDSQGIKIVTQEKELVOPFSSLPKVE--YIARAGWTRDGYAAMAFDRPQOMLOLVL 381
Db 298 ----GLNGPHTDLEMPDD--PRMREYITTYMKATSTKVA-VTMINRQANNVSIILFLC 349
QY 382 PPALEFISTENEQRLASARAVPRNVPYVVEEVTNWI--NVHDIFFYFPQSEGEDE 438
Db 350 DATGVCCTKKHEDE-----SEAMLHRQNDPEYF-----SKDGRK 383
QY 439 LCFIRANECKTGFCGLYKVTAVLKSQGYDMSPEPSGDEKCPKIEKIALTSGEMVELA 498
Db 384 FFFIRAIIP--QGGRCRKFYHIV-----SSSQPNSSNDN-----IQ---SITSGDMV-- 425
QY 499 RHGSKIWV--NEETKIVYFOGKTQPTLEHHLVYSYEAGELIVRLTTPGFSSH---SCSMQ 554
Db 426 ----TKLIAVDEKGNKIYFLSTEDLPRRROLYSANTEG-----NFNQCLSCDLVE 472

QY 555 NFDMEVSHYSVSPPCVHYVYKLSGPD-----DDPLHKQPRFASMAE--AASCPDVP 607
Db 473 NCTYFASFSHSMQ---FLLKCEGPGVPVYVNTJTDKKMKMFLDLENEHVKKAIINDROM 529
QY 608 PEIHFHTR--SDVPLYGMKYKPHALQGGKHPTVLPFYGGQVOLVNNSEFGIKYLRANT 666
Db 530 PKVEYRIEIDYVNLPMQILKPATFTDTTHYPLLVVDGTPGSSQVKEKEF-VSMETV-M 587
QY 667 LASGYAVVYIDGSGGORGRLFEQALKNQMGVEIEDQEGLOFVAEKYFIDLSRAVI 726
Db 588 VSHGAVVYKCDGSGSGFGKLLHEVRRGLLEEKDQMEAVRTML--KEYIDRTYAV 646
QY 727 HGWSYGFSLMGLIHK---POYKVAIAGAPVTYMAVDQTYTERYMDVPENNQGYE 782
Db 647 FGKDYGLSTYIIPAKGNOGQFTFCGSALSPITDFKLYASASERYLGJLHGDNDAYE 706
QY 783 AGSVALHVEKLPNPNRLIIGHFLDENVHFHTNELYSQLIRAGKPYOQIYPERHSI 842
Db 707 MTKVAHRYASAL--EEQGFLLIHPADEKIHFOHTAELITOLIRKANVSIQIYDESHYF 764
QY 843 RCPESGEHYEVTLLHFLQ 861
Db 765 TSSSLKQHLRYSIINFYE 783

RESULT 12
154331
dipeptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: I54331
R:Yokocani, N.; Doi, K.; Wenthold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a dipeptidyl aminopeptidase IV-re
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I54331
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-865 <RES>
A:Cross-references: GB:M96859; NID:g306705; PIDN:AAA35760.1; PID:g306706
C:Superfamily: dipeptidyl-peptidase IV

Query Match 9.68; Score 445.5; DB 2; Length 865;
Best Local Similarity 23.88; Pred. No. 6,4e-26;
Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;
QY 150 LFLQASNSLFCHCDGKNGFMVSPMKPLEIKTQCSGRMDPKICPADPAFFSTINSD 209
Db 255 IFIRE--NNIYYCAHVQKQAIRV----- 275
QY 210 WVAIETGEERRLTFCHQGLSNVLDLDPKSAVATFVIOEEDRFTGYWMCPTASMEGSEG 269
Db 276 ----VSTGKEGVI---YNGLSDWLYE-----EILKTHIAHWSPDGT----- 311
QY 270 LKTRILYEEDSEVEVYIHPV-----PALERKTDSTYPTGSKNPYIAKLAIEFQ 323
Db 312 ----RLAYAAINDSRVPMELPTGTSTYPTVK-----PYHYPRAGSNPISLSHVI--- 359
QY 324 TDSQGIKIVTQEKELVOPFSSLPKVE--YIARAGWTRDGYAAMAFDRPQOMLOLVL 381
Db 360 ----GLNGPHTDLEMPDD--PRMREYITTYMKATSTKVA-VTMINRQANNVSIILFLC 411
QY 382 PPALEFISTENEQRLASARAVPRNVPYVVEEVTNWI--NVHDIFFYFPQSEGEDE 438
Db 412 DATGVCCTKKHEDE-----SEAMLHRQNDPEYF-----SKDGRK 445
QY 439 LCFIRANECKTGFCGLYKVTAVLKSQGYDMSPEPSGDEKCPKIEKIALTSGEMVELA 498
Db 446 FFFIRAIIP--QGGRCRKFYHIV-----SSSQPNSSNDN-----IQ---SITSGDMV-- 487
QY 499 RHGSKIWV--NEETKIVYFOGKTQPTLEHHLVYSYEAGELIVRLTTPGFSSH---SCSMQ 554
Db 488 ----TKLIAVDEKGNKIYFLSTEDLPRRROLYSANTEG-----NFNQCLSCDLVE 534

```

QY 555 NEDMFVSHYSSVSRPCVHVYKLSGPD-----DDPLHKQPRFWMASME-AASCPDDYVP 607
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 535 NCTYSASFNSHMD---FFLTKCESGPVPMVTHNTTDKKKMFDELTENHVAKKALNDQM 591
QY 608 PEIIFEHTR-SDVRLYGMITKPHALQPGKHPTVLEFYGGPOVOLVNNSEKIKYLRILT 666
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 592 PKVEYRDIETIDYDYNLPMQILKPAFTPTDTHYPLLLVVDGTPGSGVAEKEFV-SVMEYV-M 649
QY 667 LASIGYAVVVDIGRSCGRLREFEALKNQGVLEDOVVELOVAEYKGTIDLSRAVI 726
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 650 VSSHGAIVVKKDGRSGSGFTGLLHEVRRRLGLEEKQDMEAVRTML-KEYIDSTRVAV 708
QY 727 HGWSTGCELSLMGLIHK-----POVFKAIVAGAPVTVMAYDTGTERYMDVPENNQHYE 782
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 709 FGKDVGGTLYTIIIPAKGENQGTTCGSALSPITDFKLYASAFSEKRYLGLHGLNRAVE 768
QY 783 AGSVALHVEKLPNEPNRLILHGLDENVHEFTNFLVSQILIRAGKPYQLQIYPNERHSI 842
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 769 MTKVAHVSAL--EEOQFLIHPTADEKIHFOHTAELITOLIRKANYSLOIYDPESHYF 826
QY 843 RCPESGEHYEVLHFLQE 861
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 827 TSSSLKQHLKYSIINFVE 845

```

RESULT 13

```

A11793
dipeptidyl aminopeptidase-like protein (EC 3.4.14.-) short form, DPPX-S - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A11793
R:Maeda, K.; Yokotani, N.; Hunter, C.; Dol, K.; Wenthold, R.J.; Shimasaki, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 197-201, 1992
A:Title: Differential expression of two distinct forms of mRNA encoding members of a di
A:Reference number: A11793; MUID:92108018; PMID:1729689
A:Accession: A11793
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-803 <MAD>
A:Cross-references: GB:M76429; NID:9408719; PIDN:AA041623.1; PID:9408720
A:Note: sequence extracted from NCBI Backbone (NCBIF:75138)
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; glycoprotein
F:257,342/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match 9.5%; Score 442.5; DB 2; Length 803;

Best Local Similarity 24.2%; Pred. No. 9.8e-26;

Matches 177; Conservative 108; Mismatches 287; Indels 159; Gaps 28;

```

QY 150 LFLPQASLSLFRGRCGKNGFVSMKPLEIKTCGSGPRMDEKICPADPAFFSFINSDL 209
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 193 IFIFE--NNIYYCAHVGAQAIRV----- 213
QY 210 WVANIEETGEERLTFCHQGLSNVLDPKSAGVATFVIOEEDFRFTGYMWCPTASWEGS 269
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 214 ----VSTGKEGVI---YNGLSDWLIE-----EILKTHIAHMSDPGT----- 249
QY 270 LKTLRILEYEVDSEVEVIHVPSPALERT-DSYRYPRTGSKNKRIALIAEFQTDG 328
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 250 ----RLAATINDSRVPMELPTTGYVPTAKPYHACGENSISLHVI-----G 298
QY 329 KIVSQKEIIVOPFSSLPKVE--YIARAGWTRDGIYAMAMLDLPQOQLVLPALF 386
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 299 LMGPHDLEMTTPDD--PRMREYITIMVKMATSTKVA-VNLSRAQNVSTLTLCDATG 354
QY 387 IPSTNEQORLASARAVRNVOYVYVEVTNWI---NVHDIFFPPOSSEGEDELFLR 443
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 355 VCTKHHEE-----SEAMHLRQNEEVF-----SKGRKFFVFR 388
QY 444 ANECTGTCGLHYKAVALKSGQDMSPPSPGEDEKCIKEEIALTSGEENVLARHGSK 503
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 389 AIP--OGGGGKFFYHIV-----SSSOPNSNDN-----IQ---SITSGDMV-----TK 427

```

```

QY 504 IW-VNEETKLYVYOGKTDPLEHHLYVYSFAAGELIVRLTPGFSHSCSMQNFDFVSH 562
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 428 ILSYDEKSKQIYFTLEDLPRRQLXSAS-----TVSFFNRQCLSDLYNCTYF 477
QY 553 YSSVSTPPCVHVYKLSGPD-----DDPLHKQPRFWMASME-AASCPDDYVPEI-FRHH 614
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 478 SASFSPGADFFLKCEGCPVPTVSVHNTTDKKKMFDELTENHVAKAISDRQMPKVEYRI 537
QY 615 TNSDVRKGMITKPHALQPGKHPTVLEFYGGPOVOLVNNSEKIKYLRILTASIGAV 674
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 538 ETDIDNPLQLIKPAFTPTDTHYPLLLVVDGTPGSGVAEKEFV-AATWETV-VYSSHGAVV 595
QY 675 VVIDRSGCQRLREFEALKNQGVLEDOVVELOVAEYKGTIDLSRAVIHGSWYGEF 734
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 596 VACDGRSGSGFTGLLHEVRRRLGLEEKQDMEAVR-VMLKEPYIDKTRVAVFGKDYGGY 654
QY 735 LSLMGLIHK-----POVFKAIVAGAPVTVMAYDTGTERYMDVPENNQHYEAGSVALHV 790
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 655 LSTYLLPKAGDQAVVFGSGSALSPITDFKLYASAFSEKRYLGLHGLNRAVEMAKVAHRY 714
QY 791 EKLPNENRLLILHGLDENVHEFTNFLVSQILIRAGKPYQLQIYPNERHSIRCPESGEH 850
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 715 SAL--EGQOFLVIAHTADEKIHFOHTAELITOLIRKANYSLOIYDPESHYSSAALQGH 772
QY 851 YEVTLLHFLQE 861
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 773 LHRSLTGFEVE 783

```

RESULT 14

A30107

```

dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR028C
C:Species: Saccharomyces cerevisiae
C:Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 23-Mar-2001
C:Accession: S46780; A30107
R:Du, Z.

```

submitted to the EMBL Data Library, June 1994

A:Description: The sequence of S. cerevisiae cosmid 8082.

A:Reference number: S46773

A:Accession: S46780

A:Molecule type: DNA

A:Residues: 1-818 <DU>

A:Cross-references: EMBL:U10399; NID:9500689; PIDN:AA86879.1; PID:9500698; GSPDB:GNO

J.Roberts, C.J.; Pohlig, G.; Rothman, J.H.; Stevens, T.H.

J. Cell Biol. 108, 1363-1373, 1989

A:Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an

A:Reference number: A30107; MUID:89174971; PMID:2647766

A:Accession: A30107

A:Molecule type: DNA

A:Residues: 1-82, 'H', 84-124, 'N', 126-181, 'RRLET', 189-199, 'N', 201-365, 'DFKRGKRRK', 376

A:Cross-references: EMBL:X15484

A:Note: the authors translated the codon ACC for residue 572 as Asn

C:Genetics:

A:Gene: SGD:DAF2; STE13; MIPS:YHR028C

A:Cross-references: SGD:S0001070; MIPS:YHR028C

A:Map position: 8R

C:Superfamily: dipeptidyl-peptidase IV

C:Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast v

F:30-45/Domain: transmembrane #status predicted <TM>

F:63,79,110,139,392,421/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.3%; Score 432.5; DB 1; Length 818;

Best Local Similarity 25.4%; Pred. No. 6e-25;

Matches 164; Conservative 91; Mismatches 262; Indels 129; Gaps 28;

```

QY 240 GATFVIOE-PDRFTGYMWCPTASWEGSGKTLRILEYEVDSEVEVIHVPSPALERT 298
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 234 GKPDVYEEVEEDKAAAMWSPTGD-----LAFKIDSESVGFTIIPYVQDEK 283
QY 299 ----KTDYRYPRTGSKNKRIALIAEFQTDGKIVSQKEIIVOPFSSLPKVE----- 350
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

Db 284 DIPEPMSIKYPSKSPNPHAEIWMY SMDG-----TSFHPRI SGNKK 326
 Oy 351 ----YIARAGWTRDGGYAAAMALDRPQOHLQVLLPRLAFLFISTEEDEROLASARA VPRN 406
 Db 327 DGSLLTEVTWVGNGN-VLVKTTDRSSDLITVFLIDT---IAKTSN----- 368
 Oy 407 VQPYVVEEVTN--VINVNHDFIYPPQSEGEDELCEFLRANC-----KTGFCHLKTVAV 460
 Db 369 ----VYRNSSSNGGWMEITHNTL-----FIANETFDPRPHNGT-----VDI 405
 Oy 461 LKSQGYDWESEPF-SPEDEDFKCPIKEALITSGEMEWLARHSGKIWNBEETLVYFOGTR 519
 Db 406 LPIGGYNHLAYFENSNSHYK-----LTLEGKMEVY--NGPLAFDSMENRL-YFISTR 455
 Oy 520 DIPLEHNLVSYEAGELVRLTTPG---FHSCHMSQNFMEVSHSVSTPPCV--H 573
 Db 456 KSTERHHVYIDLRSPNEIIEVTLTDSSEGDVYDVSPSSGRFGILLYKGPVYOKIIVPFH 515
 Oy 574 VKLSGPRDDPLHKORFEMASMEEA-ASCPDPY-VDPPEF-----HFHRSQVRLYG 623
 Db 516 SRKAEKCDGNVLGSKLYLLEKNEVLTKLEDYAVPRKSFRELNGKDEFGDILVNSTE 575
 Oy 624 MIYKPHALOPGKHHPTLVFVGGPQVOLVNNSEK-GIKYLRINTLASTLGAAVVIDRGSS 682
 Db 576 ILPNDDELTSDHYRYVFFAYAGGPNQQVYKTFESVGFNEV---VASQLNATVYVVDGRGT 632
 Oy 683 CORGLFESGALKNQMGVEIEDQVGLQFAEKYG--FIDLSRVAIHGWSYGGELSLMG 739
 Db 633 GPKGQDFRSLVDRDLQDYEARDOIS---AASLYGSLTFVDPQKISLFGMSYGYLTILKT 688
 Oy 740 LIHK-POVFKVVALAGAPVYVMAAYDGTGYTERYMDVPENNOHCEAGSVLAHYEKLPENN 798
 Db 689 LEKDGGRHKYKGSVAPVDMRFYDSYTERYMHPTQENFDGYVSSV--HNVTALAQAAN 746
 Oy 799 RLILHGFLENNVHFHT-NFLVSLIRAGKPYQLQIYENRHSR 843
 Db 747 RFLMHGTGDDNVHQNLSLFLDLDLNGEVNTDVHVFPSDSHSR 792
 RESULT 15
 A49737
 dipeptidyl aminopeptidase (EC 3.4.14.-) - Yeast (Saccharomyces cerevisiae)
 Alternative names: protein OS045; protein YOR219c; protein YOS50-9
 C:Species: Saccharomyces cerevisiae
 C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 29-Oct-1999
 C:Accession: A49737; S45451; S60946; S67112; S71721
 R:Santa Anna-A, S.; Herskowitz, I.
 submitted to the Protein Sequence Database, July 1993
 A:Reference number: A49737
 A:Accession: A49737
 A:Molecule type: DNA
 A:Residues: 1-931 <SAN>
 A:Cross-references: GB:L21944; NID:9347196; PIDN:AAA35119.1; PID:9347197
 R:Anna-Arriola, S.S.; Herskowitz, I.
 Yeast 10, 801-810, 1994
 A:Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase
 A:Reference number: S45451; MUID:9506382; PMID:7975897
 A:Molecule type: DNA
 A:Residues: 1-931 <ANN>
 A:Cross-references: EMBL:L21944; NID:9347196; PIDN:AAA35119.1; PID:9347197
 R:Gallison, F.; Dujon, B.
 submitted to the EMBL Data Library, October 1995
 A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome
 A:Reference number: S60938
 A:Accession: S60946
 A:Molecule type: DNA
 A:Residues: 1-931 <GAL>
 A:Cross-references: EMBL:X92441; NID:91050762; PIDN:CA63182.1; PID:91050771
 R:Boyer, J.; Fairhead, C.; Gallison, L.; Gallison, F.; Michaux, G.; Thierry, A.; Dujon, B.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67104
 A:Accession: S67112

QY 811 VHFHTNPLVSQIRAG-KPYOLOIYPNERHSIR 843
| | : | | | | : | : | |
Db 866 VHIQNTFRLVDQNLGLTNYDMHIFPDSDHSIR 899

Search completed: December 12, 2002, 12:06:37
Job time : 29 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 11:54:42 : Search time 14 Seconds
(without alignments)
2556.719 Million cell updates/sec

Title: US-09-976-674-3
Perfect score: 4646
Sequence: 1 MATTGPTADRGDAATDDP.....CPESGEHYETLLHFLQGYL 863

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	505.5	10.9	765	1	DPD4_BOVIN
2	498.5	10.7	767	1	DPD4_RAT
3	489.5	10.5	766	1	DPD4_HUMAN
4	486	10.5	765	1	DPD4_FELCA
5	481	10.4	760	1	DPD4_MOUSE
6	447.5	9.6	859	1	DPD6_RAT
7	445.5	9.6	865	1	DPD6_HUMAN
8	442.5	9.5	863	1	DPD6_BOVIN
9	438.5	9.4	761	1	SEPR_MOUSE
10	432.5	9.3	818	1	DAP2_YEAST
11	426	9.2	760	1	SEPR_HUMAN
12	417	9.0	931	1	DAP1_YEAST
13	249	5.4	657	1	YUHL_BACSD
14	215.5	4.6	732	1	ACPH_HUMAN
15	212.5	4.6	732	1	ACPH_RAT
16	194.5	4.2	732	1	ACPH_PIG
17	191	4.1	761	1	YL31_CAREL
18	165.5	3.6	705	1	PPCE_FLAME
19	160	3.4	710	1	PPCE_BOVIN
20	154	3.3	710	1	PPCE_PIG
21	151.5	3.3	705	1	PPCE_FLAME
22	144	3.1	710	1	PPCE_MOUSE
23	138.5	3.0	689	1	PPCE_AERHY
24	136	2.9	710	1	PPCE_HUMAN
25	121.5	2.6	1427	1	ABCL_SCHPO
26	118.5	2.6	726	1	Y4NA_RHISN
27	115	2.5	436	1	PAFA_CAVPO
28	115	2.5	690	1	PTRB_MORLA
29	113.5	2.4	717	1	MCCA_MOUSE
30	113	2.4	857	1	LOX3_SOYBN
31	112.5	2.4	861	1	LOX3_PEA
32	112	2.4	557	1	SASB_ANAPL
33	110.5	2.4	550	1	CRYS_DICDI

34	110.5	2.4	1246	1	VP03_HSVSA
35	108	2.3	4351	1	PAT2_RAT
36	107.5	2.3	422	1	PAFA_CHICK
37	106.5	2.3	444	1	PAFA_BOVIN
38	105	2.3	1442	1	YJ9F_YEAST
39	104.5	2.2	673	1	KPCB_BOVIN
40	104.5	2.2	722	1	Y174_RICPR
41	104	2.2	994	1	BGAL_STAXY
42	102	2.2	3144	1	VP13_YEAST
43	101.5	2.2	664	1	PD13_SHEEP
44	101	2.2	550	1	INVI_HANAN
45	100.5	2.2	635	1	TRG4_ECOLI

ALIGNMENTS

RESULT 1
DPD4_BOVIN STANDARD: PRT: 765 AA.
AC P81425.08WMC8;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26) (adenosine deaminase complexing protein) (ADCP-I)
DE (Activation molecule 3) (ACT3) (WC10).
GN DPD4 OR CD26.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.
NCBI_TaxID:9913;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-Lymphocytes;
RC PubMed:12073152;
RX Lee S.-U., Park Y.-H., Davis W.C., Hamilton M.J., Naessens J., Bonach G.A.;
RA "Molecular characterization of bovine CD26 upregulated by a staphylococcal superantigen."
RT Immunogenetics 54:216-220(2002).
[2]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RP TISSUE-Thymus;
RC PubMed:11981836;
RX Gliddon D.R., Howard C.J.;
RA "CD26 is expressed on a restricted subpopulation of dendritic cells in vivo."
RT Eur. J. Immunol. 32:1472-1481(2002).
[3]
RN SEQUENCE OF 1-24.
RP TISSUE-T-cell;
RC PubMed:11598101;
RX Lee S.-U., Ferens W., Davis W.C., Hamilton M.J., Park Y.-H., Fox L.K., Naessens J., Bonach G.A.;
RA "Identity of activation molecule 3 on superantigen-stimulated bovine cells is CD26."
RT Infect. Immun. 69:7190-7193(2001).
[4]
RN SEQUENCE OF 537-546.
RP TISSUE-Kidney;
RC MEDLINE:98293306; PubMed:9629661;
RX Ben-Shooshan I., Parola A.H.;
RA "The CP-I subunit of adenosine deaminase complexing protein from calf kidney is identical to human, mouse, and rat dipeptidyl peptidase IV."
RT Comp. Biochem. Physiol. 119B:289-292(1998).
[5]
RN FUNCTION: Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. Binds and regulates the activity of ADA.
[6]
RN CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-

```

CC      xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC      xcc is neither Pro nor hydroxyproline.
CC      -1- SUBUNIT: Homodimer.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
CC      a soluble form (by similarity).
CC      -1- TISSUE SPECIFICITY: Intestinal epithelium, dendritic cells and
CC      several immune system tissues.
CC      -1- PTM: The soluble form (SDP) derives from the membrane form (MDP)
CC      by proteolytic processing (by similarity).
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 98B.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF461806; AAL67836.1; -.
DR      EMBL; AY056834; AAL23628.1; -.
DR      MEROPS; S09.003; -.
DR      InterPro: IPR002469; DPPV_N_term.
DR      InterPro: IPR001375; Peptidase_S9.
DR      InterPro: IPR002471; Prol_endopep_ser.
DR      InterPro: IPR000379; Ser_estrs_site.
DR      Pfam; PF00930; DPPV_N_term; 1.
DR      Pfam; PF00326; Peptidase_S9; 1.
DR      PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KM      Hydrolyase; Aminopeptidase; Dipeptidase; Serine protease;
KM      Transmembrane; 1
FT      CHAIN          1
FT      DOMAIN        38      765
FT      TRANSMEM      7      29
FT      ACT_SITE      629      629
FT      ACT_SITE      707      707
FT      ACT_SITE      739      739
FT      CARBOHYD       84
FT      CARBOHYD      91
FT      CARBOHYD      149
FT      CARBOHYD      149
FT      CARBOHYD      218
FT      CARBOHYD      228
FT      CARBOHYD      271
FT      CARBOHYD      280
FT      CARBOHYD      320
FT      CARBOHYD      392
FT      CARBOHYD      495
FT      CARBOHYD      684
SQ      SEQUENCE      765 AA; 88369 MM; E32165421F43E116 CRC64;
Query Match          10.9%; Score 505.5; DB 1; Length 765;
Best Local Similarity 25.4%; Pred. No. 3,7e-30;
Matches 180; Conservative 98; Mismatches 269; Indels 163; Gaps 28;

```

```

Db      316 -----RRIQNYSIMDICDRSTGRWIS----- 338
Qy      433 SEGDELCLFLRANCKTGFCHLYKVTAVLKSQYDWSPEFSGDEF-----KCP1K 484
Db      339 -----SVGRQHEISTT-----GWGRFPAPRPHPTSDGNSYK1IISN 376
Qy      485 EE-----IALTSGEMVELARHSGKIWNNEETKLVYFOGT-KDPLLEHLLY 528
Db      377 EEGKHICHFOTDRKNCFTFKGAMVEYIG-----IEALTSPLYLYISNEYGMGPARNLY 431
Qy      529 VVSEIAGEIYRLITTPGFSHSCM-----SCNPFMFVSHYSVSTPPCVVHYK1SGP 580
Db      432 KIQLNDYTKYVCL-----SCLEINPDRCOYYSFSGQAKYVQLKRCSGDELPLYTLHNS 484
Qy      581 DDDPLHKOPFRWASMEASCPPIYVPP--EIFFPHTRSDVRLKGM1YKPHALDPRKHP 638
Db      485 NND---KELRVLNNSDLDVLDVQNPSSKLDIHLHGTFKYMQLPPH-FKSKYRP 540
Qy      639 TVLFVGGPOVQVLNNSFKIKYLRNLN---TLASLGYAVVY-IDRGSQGRLEFEGALK 694
Db      541 LLEEVYAGPCSQKADALF-----RLMWATYLASTENIYASFDRGSGYGGDKIMHAIN 594
Qy      695 NQMQQVEIEQVEGL-QVNAEKYGFIDLSRAVLIHNSYSGFSLMGLIHKQYKVALAG 753
Db      595 RRLGTFEVEDQIEATRFQF--SKMGFVVDKRLAIWGSYGGVYTMVLAGSGVGC8IAV 652
Qy      754 APVTVMAADPTGYTERYMDV--PENNOHYEAGSVLHVEKLPNEPRLILIHGFLENDV 811
Db      653 APVSKMEYIDSVYTERIMGLPTPEDNLDSTNSTRVMSRAENF--KQVEYLILHRTADNV 710
Qy      812 HFFHTNLYSOLIRAGKRYQLQIYPNRHSIRCPESGEHYEVTLHFLQE 861
Db      711 HFQSQADISKALVDAGVDFOGSMWYTDDEHGIASTAHQHYTHMSHFLQ 760
RESULT 2
DPP4_RAT
ID DPP4_RAT          STANDARD;          PRT;          767 AA.
AC      P14740;
DT      01-APR-1990 (Ref. 14, Created)
DT      01-APR-1990 (Ref. 14, Last sequence update)
DT      15-JUN-2002 (Ref. 41, Last annotation update)
DE      Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE      antigen CD26) (GPI10 glycoprotein) (Bile canaliculus domain-specific
DE      membrane glycoprotein).
GN      DPP4 OR CD26.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX      MEDLINE=89123456; PubMed=2563382;
RA      Ogata S., Mismut Y., Ikehara Y.,
RT      "Primary structure of rat liver dipeptidyl peptidase IV deduced from
RT      its cDNA and identification of the NH2-terminal signal sequence as
RT      the membrane-anchoring domain."
RL      J. Biol. Chem. 264:3596-3601(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=88068516; PubMed=3479775;
RA      Hong W., Doyle D.,
RT      "cDNA cloning for a bile canaliculus domain-specific membrane
RT      glycoprotein of rat hepatocytes."
RL      Proc. Natl. Acad. Sci. U.S.A. 84:7962-7966(1987).
RN      [3]
RP      SIGNAL-ANCHOR.
RX      MEDLINE=90338089; PubMed=1974258;
RA      Hong W., Doyle D.,
RT      "Molecular dissection of the NH2-terminal signal/anchor sequence of
RT      rat dipeptidyl peptidase IV."
RL      J. Cell Biol. 111:323-328(1990).
RN      [4]

```

RP SEQUENCE OF 281-302.
RC TISSUE=Kidney;
RX MEDLINE=94128239; PubMed=7905271;
RA Iwaki-Egawa S., Matanabe Y., Fujimoto Y.;
RT "N-terminal amino acid sequence of the 60-kDa protein of rat kidney
RT dipeptidyl peptidase IV.";
RL Biol. Chem. Hoppe-Seyler 374:973-975(1993).
CC -1- FUNCTION: REMOVES N-TERMINAL DIPEPTIDES SEQUENTIALLY FROM
CC POLYPEPTIDES HAVING UNSUBSTITUTED N-TERMINI PROVIDED THAT THE
CC PENULTIMATE RESIDUE IS PROLINE.
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
CC Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ALSO EXISTS IN
CC A SOLUBLE FORM.
CC -1- PTM: THE SOLUBLE FORM (SDP) DERIVES FROM THE MEMBRANE FORM (MDP)
CC BY PROTEOLYTIC PROCESSING.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04581; AAA41096.1; -;
DR EMBL: J02997; AAA41272.1; -;
DR PIR: A33315; A33315.
DR MEROPS: S09.003; -;
DR InterPro: IPR002469; DPPIV_N term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; Prol_endopep_ser.
DR InterPro: IPR000379; Ser_estrs_site.
DR Pfam: PF00326; Peptidase_S9; 1.
DR Pfam: PF00930; DPPIV_N term; 1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
DR Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
DR Transmembrane; Glycoprotein; Signal-anchor.
KM CHAIN 1 767
FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM
FT (MDP).
FT CHAIN 37 767
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDP).
FT CHAIN 281 767
FT DIPEPTIDYL PEPTIDASE IV 60 KDA SOLUBLE
FT FORM.
FT DOMAIN 1 6
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 28
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 767
FT ACT_SITE 631 631
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 709 709
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 741 741
FT CARBOHYD 83 83
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 90 90
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 319 319
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 521 521
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 666 666
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 38 38
FT A -> R (IN REF. 2).
FT CONFLICT 183 183
FT I -> T (IN REF. 2).
FT CONFLICT 332 332
FT T -> N (IN REF. 2).
FT CONFLICT 352 352
FT C -> V (IN REF. 2).
FT CONFLICT 394 394
FT V -> D (IN REF. 2).
FT CONFLICT 562 562
FT L -> F (IN REF. 2).
FT CONFLICT 624 624
FT R -> Q (IN REF. 2).
SQ SEQUENCE 767 AA: 88003 MW: AA17C0EC6F0E4652 CRC64:

Query Match 10.7%; Score 498.5; DB 1; Length 767;
Best Local Similarity 24.1%; Pred. No. 1.3e-29;

Matches 210; Conservative 121; Mismatches 328; Indels 213; Gaps 38;
QY 63 DESGPHSHRLY-----YLGMPIYSGRENSLLY-SEITPKVKYREKALLLSWKOMLDHPQATPH 117
Db 31 DEADA0SARLYTLADLYLKTFTFRVKSYSLRWVS0SELYLK0ENNILL-----FNA--E 80
QY 118 HGVSREELLRERRKRGVIGTISYDFH--SEGLPFL0ASNSLFFHCRD0GKNGFVSPMK 176
Db 81 HG-----NSSFLFENSFFELFGDSISDYSPPDLFVLENYVK0MRHSTTASTSYIDLN 136
QY 177 PLEIKTQCSGPRMDPKICAPDAF--FSFINSDLVANIIETGERRLTFC0GLSNVLD 235
Db 137 KR0LITEEKIPNTQ0ITMS0EGHKLAYWKN0IYKIEPRLPSHRT--ST0KEVIFEN 194
QY 236 PKAGVATVYIQDE-FDRFTGYWCTPTASWEGSEGLKTLRILEYE0DESVEYIH--VP 291
Db 195 ---GINDWYEEIEFGASALWMS0NGTF-----LAAV0PNDGVPILIEYSFY 240
QY 292 SPALERKTDYSYRYPRTGSKNPKIALKAEF0TDS0GKIYST0KEKLY0FSSLPKVEY 351
Db 241 DESLQPKYIWIPIYKAGAVNPTV--KFLVNTDSLSSTTTTIPM0ITAP-ASVTTGDH 297
QY 352 IARAGTRDQKXAMAMEFLDRPQ0ML0LVLPALFTIPSTENEB0RLASARAVPRN0PYV 411
Db 298 LCDVAMVSEDRIS-----LQWL-----RIQ0NYS 321
QY 412 V-----YEEVTNWINVHDIFIPFP0SEGDELCPLRAN0CKTGF0HLKVTAVLKS0GY 466
Db 322 VMAICDYDKTTLW-----NCPYTOEH-----IETSAT 349
QY 467 DMSEPPSPGDEFR-----CPIKEE-----IATLSGEWEIARH 500
Db 350 GMCGRPRAPPHFTSSGSSFYKIVSPDKGKIKIC0Q0DKKRPQ0CTFTTKGMEVYS-- 407
QY 501 GSKIMWNEETKLVYFOGT-KDPLRHLHYVSYEAGETIVRLTTPGFSHSCS0M0FDMF 559
Db 408 ---LEALTS0LYLYISNEKEMPG0RNLKIQ-----LT0HTNKKC-LSC0LNEB 453
QY 560 VSHYSSVTPPCVHYKLS--GP-----DDPLHK0PRFMASME 597
Db 454 RCQYVSYSLSKEAKYQ0LGRGRLPYTLRST0DKELRVL0D0NAL0K-----MQ 506
QY 598 AASCPRDYPPELTFHHTS0DVL-YGMITYK0PAL0GKKHPTVLVYGG0V0Y0LVNNS 656
Db 507 DVO-----MPSKKLDFIVLNETRFWY0MLPRH-FRKS0KYPLIDVY0G0S0KADAP 560
QY 657 KGIKYLRNLN---TLASLGAVVY-IDGRSG0CQGLRFE0GALK0M0C0VEIED0VEGL-QF 711
Db 561 -----RLWMAVYLASTENITIVASFPDGRSGY0GDKRIMAH0INKRGLTLEVED0IEAR0F 614
QY 712 VAEKTFIDLSRAVHIGWSTYGGFLSLMGLIHKP0VFKVALIAGAPVTWMAVDTGYTERVY 771
Db 615 L--KMFVDSKRAVIAWMSYGGVYTSVNL0SGSGVFKCGI0AVAPSRMEYDDVYTERVY 672
QY 772 DV--PENN0HGVSAGVALHVEKLRPNRNLILH0FLDENVYFFHTNPLVSLIAR0P 829
Db 673 GLETPEDNDHFRNSYVMSRAENE--K0VEYLLIHL0TAD0NVH0QS0A0ISKALVAD0V 730
QY 830 Y0L0IYPNRHSIRCPESGEHVEVTLHLF0E 861
Db 731 F0AMWYTTDEHGIASSTAH0IYSHMSHFL0Q 762

RESULT 3
ID DP4_HUMAN STANDARD: PRT: 766 AA.
AC P27487:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (T103) (adenosine deaminase complexing protein-2)
DE (ADABP).

GN DPP4 OR ADPC2 OR CD26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=92329551; PubMed=1352704;
 RX Misumi Y., Hayashi Y., Arakawa F., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of human dipeptidyl
 RL peptidase IV, a serine proteinase on the cell surface.";
 RL Biochim. Biophys. Acta 1131:333-336(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA MEDLINE=95012454; PubMed=7927537;
 RX Abbott C.A., Baker E., Sutherland G.R., McCaughan G.W.;
 RT "Genomic organization, exact localization, and tissue expression of
 RL the human CD26 (dipeptidyl peptidase IV) gene.";
 RN Immunogenetics 40:331-338(1994).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood;
 RA MEDLINE=92325476; PubMed=1352530;
 RX Tanaka T., Camerini D., Seed B., Torimoto Y., Dang N.H., Kameoka J.,
 RA Dahlberg H.N., Schlossman S.F., Morimoto C.;
 RT "Cloning and functional expression of the T cell activation antigen
 RL CD26.";
 RL J. Immunol. 149:481-486(1992).
 RN (4)
 RP ERRATUM.
 RX MEDLINE=93171637; PubMed=8094732;
 RA Tanaka T.;
 RL J. Immunol. 150:2090-2090(1993).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN (6)
 RP SEQUENCE OF 1-551 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=92165847; PubMed=1347043;
 RA Darmoul D., Lacasa M., Baricault L., Marguet D., Sapin C.,
 RA Trotot P., Barbat A.;
 RT "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like
 RL colon cancer cell lines HT-29 and Caco-2. Cloning of the complete
 RL human coding sequence and changes of dipeptidyl peptidase IV mRNA
 RL levels during cell differentiation.";
 RL J. Biol. Chem. 267:4824-4833(1992).
 RN (7)
 RP SEQUENCE OF 545-766 FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=91024044; PubMed=1977364;
 RA Darmoul D., Lacasa M., Chantret I., Swallow D., Trugnan G.;
 RT "Isolation of a cDNA probe for the human intestinal
 RL dipeptidyl peptidase IV and assignment of the gene locus DPP4 to
 RL chromosome 2.";
 RL Ann. Hum. Genet. 54:191-197(1990).
 RN (8)
 RP SEQUENCE OF 1-31 FROM N.A.
 RX MEDLINE=96067599; PubMed=7487939;
 RA Boehm S.K., Gum J.R. Jr., Erickson R.H., Hicks J.W., Kim Y.S.;
 RT "Human dipeptidyl peptidase IV gene promoter: tissue-specific
 RL regulation from a TATA-less GC-rich sequence characteristic of a
 RL housekeeping gene promoter.";
 RL Biochem. J. 311:835-843(1995).
 RN (9)
 RP PARTIAL SEQUENCE.
 RC TISSUE=Kidney;
 RX MEDLINE=93210468; PubMed=8096237;
 RA Morrison M.E., Vijayaradhil S., Engelstein D., Albino A.P.,

RA Houghton A.N.;
 RT "A marker for neoplastic progression of human melanocytes is a cell
 RL surface ectopeptidase.";
 RL J. Exp. Med. 177:1135-1143(1993).
 CC -i- FUNCTION: Removes N-terminal dipeptides sequentially from
 CC polypeptides having unsubstituted N-termini provided that the
 CC penultimate residue is proline. Plays a role in T cell activation.
 CC -i- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-
 CC Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
 CC Xcc is neither Pro nor hydroxyproline.
 CC -i- SUBUNIT: Homodimer or heterodimer with Seprase (FAP).
 CC -i- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
 CC a soluble form.
 CC -i- PFM: The soluble form (SDPP) derives from the membrane form (MDPP)
 CC by proteolytic processing.
 CC -i- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.
 CC -i- DATABASE: NAME=PRO; NOTE=CD guide CD26 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd26.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13735; AAB60646.1; -
 DR EMBL: U13710; AAB60646.1; JOINED.
 DR EMBL: U13711; AAB60646.1; JOINED.
 DR EMBL: U13712; AAB60646.1; JOINED.
 DR EMBL: U13713; AAB60646.1; JOINED.
 DR EMBL: U13714; AAB60646.1; JOINED.
 DR EMBL: U13715; AAB60646.1; JOINED.
 DR EMBL: U13716; AAB60646.1; JOINED.
 DR EMBL: U13717; AAB60646.1; JOINED.
 DR EMBL: U13718; AAB60646.1; JOINED.
 DR EMBL: U13719; AAB60646.1; JOINED.
 DR EMBL: U13720; AAB60646.1; JOINED.
 DR EMBL: U13721; AAB60646.1; JOINED.
 DR EMBL: U13722; AAB60646.1; JOINED.
 DR EMBL: U13723; AAB60646.1; JOINED.
 DR EMBL: U13724; AAB60646.1; JOINED.
 DR EMBL: U13725; AAB60646.1; JOINED.
 DR EMBL: U13726; AAB60646.1; JOINED.
 DR EMBL: U13727; AAB60646.1; JOINED.
 DR EMBL: U13728; AAB60646.1; JOINED.
 DR EMBL: U13729; AAB60646.1; JOINED.
 DR EMBL: U13730; AAB60646.1; JOINED.
 DR EMBL: U13731; AAB60646.1; JOINED.
 DR EMBL: U13732; AAB60646.1; JOINED.
 DR EMBL: U13733; AAB60646.1; JOINED.
 DR EMBL: U13734; AAB60646.1; JOINED.
 DR EMBL: M74777; AAB60646.1; -
 DR EMBL: BC013329; AAH13329.1; -
 DR EMBL: M80536; AAA52308.1; -
 DR EMBL: X60708; CA43118.1; -
 DR EMBL: S79876; AAB35614.1; -
 DR PIR: S24313; CDH026.
 DR MEROPS: S09.003; -
 DR GeneW: HGNC:3009; DPP4.
 DR MIM: 102720; -
 DR InterPro: IPR002469; DPP4.
 DR InterPro: IPR001375; Peptidase_S9.
 DR InterPro: IPR002471; Prol_endopep_ser.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00326; Peptidase_S9; 1.
 DR Pfam: PF00930; DPP4_N term; 1.
 DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 DR Hydrolase: Aminopeptidase; Dipeptidase; Serine protease;
 KW Transmembrane; Glycoprotein; Signal-anchor.
 FT CHAIN
 FT 1
 FT 766
 FT (MDPP).
 FT DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

FT CHAIN 39 766 DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT 1 6 (SDP).
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 766 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 630 630 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 708 708 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 740 740 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 6 6 K -> R (IN REF. 6).
FT CONFLICT 7 7 V -> I (IN REF. 1).
FT CONFLICT 437 437 S -> I (IN REF. 1).
FT CONFLICT 557 557 T -> I (IN REF. 7).
FT CONFLICT 663 663 D -> E (IN REF. 7).
SQ SEQUENCE 766 AA; 88278 MM; 5FBA42C662D6117 CRC64;

```

Query Match 10.5%; Score 489.5; DB 1; Length 766;
 Best Local Similarity 22.4%; Pred. No. 6e-29; Indels 241; Gaps 35;
 Matches 201; Conservative 124; Mismatches 330;

```

QY 47 LKVNKAPDFQVORTDESGHSHRLYYLGMPSYGRNSLAYSLEIPKRYKREALLLSWK 106
DB 26 VLKNG-----TDATADSRKTYTL-----TDLKMTYRLKLYSLRW- 62
QY 107 QMLDHFQATPHHGYVSREELL-----REKRLGVGITSYDFH-SESGFLF 153
DB 63 -----LSDHELYLKQENHLLVENAEYGNSSVFLFNSTDFEGHSHINDYSISPGQPTL 115
QY 154 QASNLFLICRGGKNGKGFVSPKPLEIKTQCSGPRMDKICPADAF-FSFTNNSDIWA 212
DB 116 LEYNVKKMRHSYASTAYDIYDLNKRQLITEERIPNNTQWVSPVGHKLAYWNNDIYK 175
QY 213 NIETGERRLTFCHQGLSNVLDDPKSAGVATFVIOE-FDRRTGYWMCPTASWESSEGLK 271
DB 176 IEPNLPYRITW--TGKDDIYN--GITDMVYEEVFSAYSAIAMSNGNGF----- 222
QY 272 TLRLIIEVDESEVYH---VSPALDEKRTDSTRYRTGSKNPKIALKAEFOTDSOG 328
DB 223 ---LAVAGFNDTEVPLIEYSFYSDESLOYPRTVAVPYKAGAVNPTV--KEFVNTDSIS 277
QY 329 KIVSTQEKELVQFSSLPFKVEYIARAGWTRDGKAYAMAFDRPQOMIQLVLLPALFIP 388
DB 278 SVTNAITSIGITAPASMLGD-HYLCDDVWATQERIS-----LQWL----- 316
QY 369 STENEORLASARAVRVNVOYV-----YEEVTWVWVNDIEYFPQSEGEDELCLFLR 443
DB 317 -----RRLONSVMDICDDYDESSGRW----- 337
QY 444 ANECKTGCHLYKYTAVALKSCGYMSEPFSPQEDE-----FK 480
DB 338 --NCLVAROHLEMTST-----GWVGRFRPSEPHFTLDGNSFYKIIISNEGYRHCYFQ 388
QY 481 CPIKEIALTSGEWEVLARHGSKIWNDETCLVYFOGTKDPTEHLHLVVSYEAG----- 536
DB 389 IDKDCITITGKTWEVIGLEAL-----TSD-----YLIYSINIEYGMROG 428
QY 537 -EIVRLTTPGFSHSCSMQNFDMFVSHYSVSTPPCVHYKL--SGP----- 580
DB 429 RNLVYKIQDSYTKVCTLCSELPERCQYYSVFSKKAAYYQLRCGSGPLPLTLHSSVND 488
QY 561 -----DDDLPHKQPRRWASMEAAASCPDVPYPEIFHFRHSVRL-YGAIYYPHALO 632
DB 489 KGLRVLENSALDK-----MLQNVQ-----MPSKKLDFILNTEKTFYQWILPPH-FD 535

```

```

QY 633 PKKHPTLVFYGCPQOVOLVNNSEFKIKYLRN---TLASLGAVV--IDGRGSCORGLR 688
DB 536 KSKKPTLLDDYAGCSCQKADIVF-----RLNATYIASHTENITVASFDFGSGCYQDK 589
QY 689 FEGALKNOGOVEIEDQVEGL-QFVAERYGFTDLSRVALHGSYGGSLMGLHKPOVF 747
DB 590 IMHAINRLGTFEVEDQLEAAROF--SKMGFVDNKRRIAMGWSYGYTSMVLSGSGVFE 647
QY 748 KVAINGAVTYWMAVDTCGYTERYMDV--PENNGHVEAGSVALHKEKLPENRLLIHLG 805
DB 648 KGVIAVAPVSREYEDSVYTERYKGLPPEDNLDHYRNSVSRANF--KQVEYLLIHLG 705
QY 806 FLDENVHFHTNFIYSOLIRAGKPYOLOIYPERHSIRCPESGEHYETLLHFLOE 861
DB 706 TADDVHHTQGSQISKALVDVGDVQAMWYDDEHGIASSTHAIHTYHMSHFIRQ 761

RESULT 4
DPP4_FELCA STANDARD: PRF: 765 AA.
AC Q9N2I7, FELCA 15-JUN-2002 (Rel. 41, Created)
AC 15-JUN-2002 (Rel. 41, Last sequence update)
AC 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation antigen CD26).
GN DPP4 OR CD26.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
CX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20094000; Pubmed=10630304;
RA Nishimura Y., Miyazawa T., Ikeda Y., Izumiya Y., Nakamura K., Sato E., Mikami T., Takahashi E.;
RT "Molecular cloning and sequencing of a cDNA encoding the feline T-cell activation antigen CD26 homologue."
RL Immunogenetics 50:366-368(1999).
CC -1- FUNCTION: Removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline (By similarity).
CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-Xcc, from a polypeptide, preferentially when Xbb is Pro, provided Xcc is neither Pro nor hydroxyproline.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in a soluble form (By similarity).
CC -1- PM: The soluble form (SDPP) derives from the membrane form (MDPP) by proteolytic processing (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL: AB023952; BAA92344.1; -.
CC MEROPS: S09.003; -.
DR InterPro: IPR002469; DPP4V_N.term.
DR InterPro: IPR001375; Peptidase_S9.
DR InterPro: IPR002471; ProL_endopep_ser.
DR InterPro: IPR00379; Ser_estrs_site.
DR Pfam: PF00930; DPP4V_N.term.1.
DR Pfam: PF00326; Peptidase_S9.1.
DR PROSITE: PS00708; PRO_ENDOPEP_SER.1.
KW Hydrolase; Aminopeptidase; Dipeptidase; Serine protease; Transmembrane; Glycoprotein; Signal-anchor.
FT CHAIN 1 765 DIPEPTIDYL PEPTIDASE IV MEMBRANE FORM

```

```

FT CHAIN 38 765 (MDPP).
FT DIPEPTIDYL PEPTIDASE IV SOLUBLE FORM
FT (SDPP) (BY SIMILARITY).
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 7 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 765 EXTRACELLULAR (POTENTIAL).
FT ACT_SITE 629 629 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 707 707 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 739 739 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 765 AA; 88213 MW; 3EFC98A228175D9 CRC64;

```

Query Match: 10.58; Score 486; DB 1; Length 765;
 Best Local Similarity 23.08; Pred. No. 1,1e-28;
 Matches 200; Conservative 135; Mismatches 331; Indels 204; Gaps 33;

```

OY 59 VOKTDESGPHSHRLYLGMYPGSRNSLLSEIPKRYKREALLLSMKQMLDHFQAPPH 118
DB 28 LNKNDAAADSRKRYTL-----TDYLNKTRFRVAFYSLRW-----VSHD 66
OY 119 GVSREBELLRERRRLGFGITSYDFHSEGLFQANSLLFHCROGKNGFWSP--- 174
DB 67 YLYKQDNNILFLFNAEYG-----NSSIFL---ENSTFDEFEHSINDYSNPDQGF 112
OY 175 -----MKPLEIKTQCSGPRNDPRICADPAF--FSFINSLDM 210
DB 113 ILLEYNYVKQMRHSYASYDIYDLNKRQLITEEKIPNNQMTWSPBCHKLAUYWKNDY 172
OY 211 VANETGEERLTCFHOGLSNVLDPKSAGAVTFVIOBE--FDRTGYWCPTJASMEGSEG 269
DB 173 VKNEPNSSHRITW--TGEENAIYN---GIADVYEEIEISAKSALWMSKGF----- 221
OY 270 LKTURLIYEEDSEVEVIH--VSPALERKTDSTYRPRGSKNPKIALKLAEPQDTS 326
DB 222 -----LAVAFNDQVPLDIYSFYSDESLSQYPMTRPIYPKAGANPTV--KLEVIKDN 274
OY 327 QGKIVSQOEKELVQPFSSLPFKVEYIARAGTRDQKATAMAFLDLPQOMLOLVLLPALF 386
DB 275 LNPMTNATVEITPP--AAMLTGDIYLDVYTWANEERIS-----LQMLRIONYSYMD 325
OY 387 IPSTNEEOQLASAVPRNQPVVYEEVTVNMINVDIYFPPOSGEDELCEFLRANE 446
DB 326 IROVNNSTGKRMISSAA-----QEHII--EMSTTGWVGRRAEPHTSDGRN---FYKIIIS 375
OY 447 CKTGECHLYKVTAVLKSGYDWSPEFSGEDEFKCPIKEETALTALSGEWEVL----- 497
DB 376 NEDDYKHICR-----FOIDKKDCFFITKGAEEVIGIALTTDY 413
OY 498 -----ARHSGKITVNEETKLYVQGTQKDRPLEHNLVVSF--EAAEIVRLT 542
DB 414 LYYISNEYKMGPGGRNLYKIQINDYTVVACL--SCGLKPERCQYVSFSKAKKYQLCECS 472
OY 543 TPGRS-----HSCSMQNDMFVSHYSSVTPPCVHYVYKLSGPDODPLHKORFNASMMEA 598
DB 473 GPGILPTLHRSSNDE--ELRVLEDNSALDKMLQEVMPKSKDPIILNETKRYQMI-- 528
OY 599 ASCPPDYVPEPEIFHFRSDVRLGMYTPRALOPGRKHPTVLEFVGSPQVQVYVNNSEFKG 658
DB 529 -----LPP--HFDI-----SKKYPILLIYVAGPQCKADALF-- 558
OY 659 IKYIRLN---TLASLGAIVVY-IDGRSGCORGLRFEGALKMQMGOVELEDOVEGL-QFVA 713
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

DB 559 ----RLNMATYLASTENITIVASFDRSGSGYGDKIMHAVNRRLCTFEVEDQIEARQF-- 612
OY 714 EKYGFIDLSRAVIAIGWSYGFGLSLMGLTKRPQVRKVALAGAPVYVMAYDGTTERYMDV 773
DB 613 SKMGFVDDKRAIAGWSYGGVYVTSVNLGAGSGVRCGAVAPVSRWEYDSVYTERYMG 672
OY 774 --PENNOHGEYAGSVALHVEKLPEPNRLLHGFLENDEHFFHTNPLVSQILRAGRPYQ 831
DB 673 PTPQNDLDYKNSIVMSRAENF--KQVEYLLIHGTADDNVHFQOASQISKALVAGVDFQ 730
OY 832 LQIYPRNHSIRCPESGEHFEVTLHPLOE 861
DB 731 AMWYTDHDIAGSPAHQIHVTHMSHFQK 760

```

```

RESULT 5
DEP4_MOUSE STANDARD; PRT; 760 AA.
AC P28843;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dipeptidyl peptidase IV (EC 3.4.14.5) (DPP IV) (T-cell activation
DE antigen CD26) (Thymocyte-activating molecule) (THAM).
GN DPP4 OR CD26.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Thymus;
RX MEDLINE=92129288; PubMed=1370813;
RA Marguet D.A., Bernard A.-M., Vivier I., Darmoul D., Naquet P.,
RA Pierres M.;
RT "cDNA cloning for mouse thymocyte-activating molecule. A
RT multifunctional ecto-dipeptidyl peptidase IV (CD26) included in a
RT subgroup of serine proteases."
RJ J. Biol. Chem. 267:2200-2208(1992).
RN [2]
RP REVISIONS.
RA Marguet D.A.;
RA Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.A; TISSUE=Liver;
RX MEDLINE=95092780; PubMed=7999781;
RA Bernard A.-M., Mattei M.-G., Pierres M., Marguet D.;
RT "Structure of the mouse dipeptidyl peptidase IV (CD26) gene."
RJ Biochemistry 33:15204-15214(1994).
RN [4]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=91302787; PubMed=1712807;
RA Vivier I., Marguet D.A., Naquet P., Bonicel J., Black D., Li C.X.-Y.,
RA Bernard A.-M., Gorvel J.-P., Pierres M.;
RT "Evidence that thymocyte-activating molecule is mouse CD26
RT (dipeptidyl peptidase IV)."
RJ Immunol. 147:447-454(1991).
OY 1- FUNCTION: Removes N-terminal dipeptides sequentially from
OY polypeptides having unsubstituted N-termini provided that the
OY penultimate residue is proline.
OY CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-I-
OY Xcc, from a polypeptide, preferentially when Xbb is Pro, provided
OY Xcc is neither Pro nor hydroxyproline.
OY 1- SUBUNIT: Homodimer.
OY 1- SUBCELLULAR LOCATION: Type II membrane protein. Also exists in
OY a soluble form.
OY 1- PTM: The soluble form (SDPP) derives from the membrane form (MDPP)
OY by proteolytic processing.

```


[illegible]

ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M96859; AAA35760.1; -
 CC EMBL: M96860; AAA35761.1; -
 CC MEROPS: S09.973; -
 CC GeneW: HGNC:3010; DPP6.
 CC MIM: 126141; -
 CC InterPro: IPR002469; DPPIV_N_term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC Pfam: PF00326; Peptidase_S9; 1.
 CC Pfam: PF00930; DPPIV_N_term; 1.
 CC Transmembrane; Glycoprotein; Signal-anchor; Alternative splicing.
 CC DOMAIN 1 95 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 117 865 EXTRACELLULAR (POTENTIAL).
 CC CARBOHD 173 173 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 319 319 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 404 404 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 471 471 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 535 535 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 566 566 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC CARBOHD 813 813 N-LINKED (GLCNAc. . .) (POTENTIAL).
 CC VARSPLIC 1 81 MASLYQRTGKINTSRPAPPEASHLLGGCGPEDGAGGA
 CC KPLGRQAQAAAPRERGCGGAGGAPRFOYOGRSQDGED
 CC -> MTTAKEPSASGSKVQOQOGE (IN ISOBORN DPPX-
 CC S).
 SO SEQUENCE 865 AA; 97588 MW; 1AB1AE0024464B CRC64;
 Query Match 9.6%; Score 445.5; DB 1; Length 865;
 Best Local Similarity 23.8%; Pred. No. 1.5e-25;
 Matches 176; Conservative 120; Mismatches 268; Indels 175; Gaps 31;

DB 535 NOTYASASHSHMD---FFLKCEGPVPMVTWHTTDKKKFFDEJTNHWKKAINDQM 591
 OY 608 PEIHFHTR-SDVRLYGMITYPHALQPGKHPTVLFVYGGPOVOLVNSFKIKYLRLT 666
 DB 592 PKVEYRDEIDIDYNNPMQILKPATFTDTHYPLLVGDTPPSGSVAEKEF-VSMETV-M 649
 OY 667 LASLGYAVVVDGRSCQRLGFEAGAKNOMQVEIEQVDELQVAEKYGTIDISRAI 726
 DB 650 VSSHAAVAVVCKDGRSGFGTKLLEHVRRLGLLEKQDMENVKRL-KEQYIDRTFAV 708
 OY 727 HWSYGFSLSLGLIHK-----POVEKVALAGAPVTWMAAYDTGYERWMDVPENNQGYE 782
 DB 709 FCKDYGCIYSTIILPAKGNOCQFTCGSALSPITDFKIYASFSERYIGLGLDNRAVE 768
 DB 769 MKVAVHVSAL--EEQQLIHPTADEKIHQHTAELLITOLIRKANVSLQIYDPESHVF 826
 OY 843 RCPSEGEHVEYLLHFLQE 861
 DB 827 TSSSLKQHLNRSITNFEVE 845
 RESULT 8
 DPP6_BOVIN STANDARD; PRT; 863 AA.
 ID DPP6_BOVIN
 AC P42659;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dipeptidyl peptidase IV like protein (Dipeptidyl aminopeptidase-
 DE related protein) (Dipeptidylpeptidase VI) (DPPX).
 OS Dppe.
 GN Bos taurus (Bovine)..
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=92108018; PubMed=1729689;
 RA Wade K., Yokotani N., Hunter C., Doi K., Wenthold R.J., Shimasaki S.;
 RT "Differential expression of two distinct forms of mRNA encoding
 RT members of a dipeptidyl aminopeptidase family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:197-201(1992).
 CC -!- FUNCTION: MAY BE INVOLVED IN THE PHYSIOLOGICAL PROCESSES OF BRAIN
 CC ACTIVITY. HAS NO DIPEPTIDYL AMINOPEPTIDASE ACTIVITY. THE LACK OF
 CC ACTIVITY MAY BE DUE TO THE SUBSTITUTION OF AN ASPARTATE RESIDUE
 CC FOR THE SERINE RESIDUE IN THE PROPOSED CATALYTIC TRIAD.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: DPPX-L (SHOWN HERE) AND DPPX-S;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE BRAIN. DPPX-L
 CC IS EXPRESSED EXCLUSIVELY IN THE BRAIN WHEREAS DPPX-S IS FOUND IN
 CC BRAIN, KIDNEY, OVARY AND TESTIS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M76428; AAC41622.1; -
 CC EMBL: M76429; AAC41623.1; -
 CC MEROPS: S09.973; -
 CC InterPro: IPR002469; DPPIV_N_term.
 CC InterPro: IPR001375; Peptidase_S9.
 CC InterPro: IPR000379; Ser_estrs_site.

FT	ACT SITE	734	734	CHANGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHD	49	49	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	92	92	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	227	227	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	314	314	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD	679	679	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLC	31	35	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLC	31	63	MISSING (IN ISOFORM 2).
FT	CONFLICT	737	737	MISSING (IN ISOFORM 3).
SO	SEQUENCE	761 AA;	87944 MW;	S -> L (IN REF. 2).

Query Match	9.48;	Score 438.5;	DB 1;	Length 761;
Best Local Similarity	22.28;	Pred. No. 4.1e-25;		
Matches 183;	Conservative 126;	Mismatches 287;	Indels 229;	Gaps 35

Oy	118	HGVYSREBELLERKRRLCVFGTYSVF-----HSSGGLFLQASNSLPHC8D	16
Db	79	YNIEFRRESTIILSNSTKMSVNA7DGLSPDRQFVLESDISKLMRYSTATTYYTDLOHG	13
Oy	166	GKNGFVSPMKRPLEIKTQCSGRPMQKIC-----PADPAFFSFINNSDIWVAN	21
Db	139	---EFVRGYELPRPIQYLCWSP-VQSKLAVYYQNNITLKRGRGDPF-----QI	18
Oy	214	IETGERKULTRCHQGLSNVLDDPKAGAVATFYIOEFDRFICY-WKCP7ASWESSEGLK	27
Db	184	TYTGRENRI-----FNGJPDWYEEEM-LATKYALWMSPDGKF-----	22
Oy	272	TLRIIYEEUDESEVEVI-----HVSPALAEKRTDYSRYPR7SGKNPKATLALAEQ	32
Db	221	---LATYEFNOSDPII1AYSYGDOQY-----RTINIPRPKAGAKNPVRY-----	26
Oy	324	7DQSGKIVSTQEKELVOPRPSLEPRVEYIARAGWRDCKYAMAMELDRPQ---QWLOLVL	38
Db	265	---FIYDTTYRPHHVGWME-VPRPEMIA-----SSDYF5SWL7WSSERVOLOMLKRVQ	31
Oy	381	LPRALFISTEBEERLASARAVRPNQOYUYVEEVTNWIMVHDIIFYPRQSEBDELC	44
Db	314	NVSVLSTICDFREDWH---AMECPKN-OEHW-EBSRTGMAGC---FVY5TPAFSODATS	36
Oy	441	FLRANECKTGFCHLKATYAVLKSQGYDMSEPRSPGDEKCPCKEKEIALTSGEVEYLARH	50
Db	364	YKKIFSDMDQYKHNIYI-----KQIVENAIQITTSKMEAI---	39
Oy	501	GSKIWNBEETKLYV---POGTKDP7LEHNL7VVSYEAGETVRL7TPGF5HSGM---	55
Db	399	---YI7RV7ODSLF5YSSNEFEG---YPRGRNIRI5IGNSP5KCVT-----CHLKER	44
Oy	553	---SQNDFMF5SH5SVSTPRCCVNYKL-SQPDDDLH-----KQPR---	59
Db	448	COYTASFSYAKKATYALCYGPGDL7ISTLHD8RTQEOI0VLEENKLELNSLRNIDPKYE	50
Oy	591	-----FWASMMEAASCPDPVPELIFHTFRSD7RLGMIYKPHALOPKRNHPV	64
Db	508	IKKLMDGGLTFWYKMI-----LPQq-----PDRSKYRPLL	53
Oy	641	LFVYGGPOVQULYNN5F-KGIKYLRLN7L7ASIGVAVVYIDGSGCQGRG7REFGALKQMG	69
Db	538	IOVYGGPC5OSVYSVFAWNMIYYL---ASKGIVAI7ALVDG7GTAFOGDKFLHAYRYKLG	59
Oy	699	QVEIDDOVEGLOFVNEKGEFIDLSVVAL7HGM5YGGFL5MGJ1HRPOYFKVIAAGAPYV	75
Db	594	YVEVDQ7L7ANAKF7E-MGFIDEETAL7MG5YGGV5SL7ALASGTGLFKC7IAA7PVS	65
Oy	759	WMA7DTG7ERYMDPV-ENNOHG7EAG5VAL7HVEK7LEPNRRL7ILHG7FIDEN7HFHT	81
Db	653	MEY7ASIV5EREMG7PTKODN7LEHYKNS7VMA7REY7FRVVD---YLL7GSTADNVH7FQMS	71
Oy	817	NFLV5QLIRACK7PYQLO7IY7PNERHS7ICP7ESE7EH7EVL7LH7LOE	861
Db	711	AOIARALVAQVDFQAMWYSDQNHQISSGR5ONHLYTHN7H7LKQ	755

RESULT	10
DAP2_YEAST	
ID	DAP2_YEAST
STANDARD:	PRT; 818 AA.
AC	PI8962;
DT	01-NOV-1990 (Rel. 16, Created)
DT	01-FEB-1995 (Rel. 31, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Dipeptidyl aminopeptidase B (EC 3.4.14.-) (DPAB B) (YSCV).
GN	DAP2 OR YHR028C.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX	NCBI_TaxID=4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89174971; PubMed=2647766;
RX	Robertes C.J., Pohlig G., Rothman J.H., Stevens T.H.;
RA	"Structure, biosynthesis, and localization of dipeptidyl
RT	aminopeptidase B, an integral membrane glycoprotein of the yeast
RT	vacuole."
RL	J. Cell Biol. 108:1363-1373(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=S288c / AB972;
RX	MEDLINE=94378003; PubMed=8091229;
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA	Du Z., Favella A., Fulton L., Gatlung S., Geisel C., Kirsten J.,
RA	Kucaba T., Hillier L., Jler M., Johnston L., Langston Y.,
RA	Lafreille P., Louis E.J., Macri C., Maris E., Meneses S., Mouser L.,
RA	Nhan M., Rifkin L., Riley L., St Peter H., Treaskis E., Vaughn K.,
RA	Vinuet D., Wilcox L., Woldman P., Waterston R., Wilson R.,
VA	Vaudin M.;
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT	VIII."
RL	Science 265:2077-2082(1994).
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. LYXOSOME-LIKE
CC	VACUOLES.
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X15484; CA335512.1; -;
DR	EMBL; U10399; AAB68879.1; -;
DR	PIR; A30107; A30107.
DR	PIR; S46780; S46780.
DR	MEROPS; S09_006; -;
DR	SGD; S0001070; DAP2.
DR	InterPro; IPR002469; DPPIV_N.term.
DR	InterPro; IPR001375; Peptidase_S9.
DR	InterPro; IPR002471; ProL_endop_ser.
DR	InterPro; IPR000379; Ser_estr_site.
DR	Pfam; PF00326; Peptidase_S9; 1.
DR	Pfam; PF00930; DPPIV_N.term; 1.
DR	PROSITE; PS00708; PRO_ENDOPEP_SER_1.
RW	Hydrolase; Aminopeptidase; Dipeptidase; Serine protease;
KW	Transmembrane; Glycoprotein; Signal-anchor.
FT	DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 30 45 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	DOMAIN 46 818 LUMENAL (POTENTIAL).
FT	ACT_SITE 679 679 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 756 756 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 789 789 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 738 738 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 83 83 O -> H (IN REF. 1).
FT CONFLICT 125 125 S -> N (IN REF. 1).
FT CONFLICT 182 188 FEELING -> LIRLET (IN REF. 1).
FT CONFLICT 200 200 D -> N (IN REF. 1).
FT CONFLICT 366 375 TSNVNRSS -> DEKGERKE (IN REF. 1).
FT CONFLICT 808 818 AKRAFDOQEVK -> OSVLSMGMLTIELIYSSHRDHKT
FSYLTMYI (IN REF. 1).
SQ SEQUENCE 818 AA: 93404 MW: 318f45045375BD3 CRC64;

Query Match 9.38; Score 432.5; DB 1; Length 818;
Best Local Similarity 25.4%; Pred. No. 1.3e-24;
Matches 164; Conservative 91; Mismatches 262; Indels 129; Gaps 28;

OY 240 GVAFVIOEE-FDRFGYWCPTASWEGSEGLKRLILEYVDESEVAVHVPALFEER 298
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
D 234 GKPMVVEEVEFEDDKAAMWSPTGDT-----IAFLKIDSEGEFTIPIYVODEK 283
OY 299 ----KTDYRYPRTGSKNPRIALKLAEFQTDGKIVSTOEKELVQFSSLPKVE---- 350
D 284 DIYEMHSIKYPSKGTNPNAELWVSMKG-----TSFHPRISSGK 326
OY 351 ----YIARAGTDRDKYAMAFELDRPOOMLOLVLPALPITSTENEORLASARAVPRN 406
D 327 DGSLLILEVAVWNGN-VLVKTDRSSDILTVFLIDT---IAKTSN----- 368
OY 407 VQPYVVEEYTN-VWIVNDHIEFPPOSSEDELCFLRANEC---KTGRCHLYKTAV 460
D 369 ----VWNESSNGWMTITNTL-----FIPANEFTDRPHNGT-----VDI 405
OY 461 LKSGYDWSPEF-SPEGEDEFKCPKEIEIALTSGEMEVLARHGSKIVNEETKILVFOGTR 519
D 406 LPIGVYHLAVFEMSNSHK-----TLDEGKEVY-NGPLAFISMENRL-YFISTR 455
OY 520 DTPLEHLLVYVSEAGEIYRLTPG---FSHSCSMQNDMFVSHYSVSTPCV--H 573
D 456 KSSIERHVVYIDLRSPNEIIEVTDSEDGYDVFSFGRRFGLLTYKGPVPYOKIYDFH 515
OY 574 VYKLSGPDOPRLHKOPRFMA5MMEA-ASCPPDY-VPEEIF-----HFHTSDVRLYG 623
D 516 SRKAERKDKGNVLGSKYLHEKNEVLTILEDYAVPRKSRRELNLGDEGKDLVNSTE 575
OY 624 MIYKRALDGRKHPYLVFYGGQVQVQVYVNNSEK-GIKYLRNLTLASGYAVVYIDGRGS 682
D 576 ILRPDEFETLSHDYRPFYFAYGPNQGVKTFSGFENEY---VASQLNALIVVVDGRGT 632
OY 683 CQGLRREGALKNOMGVLEIDYEGIQFYAEKYG---FIDLSVNAIHGMSYGGFSLMG 739
D 633 GFGQDQRLSVRRLDGYEARDQIS---AASLYGSLTFVDPKISLFGMSYGGYLTLLKT 688
OY 740 LIHK-POVFVVALAGAVTWMAVDTGTYERYMDVPRNNHNGYAGSVALHVEKLPRPN 798
D 689 LEADGGHFKYGSVAVPTWMPRFDSYTERYMTPEANDGYESSV--HNVTALAQAN 746
OY 799 RLILHGLDENVHFHT-NFIVSQLIRACKPYQLOIYPERHSIR 843
D 747 RFLIMHGTGDNDVHFQNSLKEFLDLLDNGVENVDVHFPSDSHSIR 792

RESULT 11
SEPR_HUMAN
AC 012884; 099998; 000199; 090194; PRT; 760 AA.
ID 15-JUN-2002 (Rel. 41) Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Seprase (EC 3.4.21.-) (Fibroblast activation protein alpha) (Integral
DE membrane serine protease) (170-kDa melanoma membrane-bound
DE gelatinase).
GN FAP.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Fibroblast;
RX MEDLINE=94261645; PubMed=7911242;
RA Scanlan M.J., Raj B.K.M., Calvo B., Garin-Chesa P., Sanz-Moncali M.P.,
RA Healey J.H., Old L.J., Rettig W.J.;
RT "Molecular cloning of fibroblast activation protein alpha, a member of
RT the serine protease family selectively expressed in stromal
RT fibroblasts of epithelial cancers."
RT Proc. Natl. Acad. Sci. U.S.A. 91:5657-5661(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Melanoma;
RX MEDLINE=97388251; PubMed=9247085;
RA Goldstein L.A., Gherzi G., Pineiro-Sanchez M.L., Salamone M., Yeh Y.,
RA Flessate D., Chen W.-T.;
RT "Molecular cloning of seprase: a serine integral membrane protease
RT from human melanoma."
RL Biochim. Biophys. Acta 1361:11-19(1997).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND SEQUENCE OF 220-229; 461-472 AND
RP 511-518.
RC TISSUE-Melanoma;
RX MEDLINE=97218181; PubMed=9065413;
RA Pineiro-Sanchez M.L., Goldstein L.A., Dotti J., Howard L., Yeh Y.,
RA Chen W.-T.;
RT "Identification of the 170-kDa melanoma membrane-bound gelatinase
RT (seprase) as a serine integral membrane protease."
RL J. Biol. Chem. 272:7595-7601(1997).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Melanoma;
RX MEDLINE=20112818; PubMed=10644713;
RA Goldstein L.A., Chen W.-T.;
RT "Identification of an alternatively spliced seprase mRNA that encodes
RT a novel intracellular isoform."
RL J. Biol. Chem. 275:2554-2559(2000).
RN [5]
RP SEQUENCE OF 192-208; 220-240 AND 510-521.
RX MEDLINE=94327249; PubMed=7519584;
RA Rettig W.J., Su S.L., Fortunato S.R., Scanlan M.J., Raj B.K.M.,
RA Garin-Chesa P., Healey J.H., Old L.J.;
RT "Fibroblast activation protein: purification, epitope mapping and
RT induction by growth factors."
RL Int. J. Cancer 58:385-392(1994).
CC -1- FUNCTION: May have a role in tissue remodeling during development
CC and wound healing, and may contribute to invasiveness in malignant
CC cancers.
CC -1- CATALYTIC ACTIVITY: Degrades gelatin and heat-denatured type I and
CC type IV collagen, but not native type I or type IV collagen. Does
CC not cleave laminin, fibronectin, fibrin or casein.
CC -1- SUBUNIT: Homodimer, or heterodimer with DPP4. The monomer is
CC inactive.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Found in cell
CC surface lamellipodia, invadopodia and on shed vesicles.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/1 (shown here) and
CC 2/s/truncated; are produced by alternative splicing. Isoform 1
CC predominates.
CC -1- TISSUE SPECIFICITY: Fibroblast-specific.
CC -1- INDUCTION: In fibroblasts at times and sites of tissue remodeling
CC during development, tissue repair, and carcinogenesis.
CC -1- PTM: N-glycosylated.
CC -1- PTM: The N-terminus may be blocked.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 12, 2002, 12:02:02 ; Search time 31 Seconds

(without alignments)

5736.086 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATGTGPTADRGDAATDDP.....CPESGHEVETLLHFLQEYL 863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	4646	100.0	863	4	Q8WXD8		Q8wx88 homo sapien
2	2887.5	62.2	883	11	Q9D4G6		Q9d4g6 mus musculu
3	2870	61.8	882	4	Q9HBM5		Q9hbm5 homo sapien
4	2572	55.4	508	4	Q75273		Q75273 homo sapien
5	2383	51.3	439	4	Q9BVR3		Q9bvr3 homo sapien
6	2137.5	46.0	632	4	Q96JX1		Q96jx1 homo sapien
7	1859.5	40.0	432	4	Q75868		Q75868 homo sapien
8	1675	36.1	312	4	Q96N78		Q96nt8 homo sapien
9	1628.5	35.1	1042	5	Q9VC20		Q9vc20 drosophila
10	1628.5	35.1	1102	5	Q9VC19		Q9vc19 drosophila
11	1494.5	32.2	465	4	Q9HBM3		Q9hbm3 homo sapien
12	1376	29.6	469	4	Q9NXP4		Q9nxf4 homo sapien
13	1220.5	26.3	360	4	Q9HEM2		Q9hbm2 homo sapien
14	1093.5	23.5	310	4	Q9HBM4		Q9hbm4 homo sapien
15	943.5	20.3	927	5	Q965K3		Q965k3 caenorhabdi
16	934.5	20.1	931	5	O44987		O44987 caenorhabdi

17	827	17.8	746	10	Q9FNF6	Q9fnf6 arabidopsis
18	805.5	17.3	738	16	Q9A6E0	Q9a6e0 caulobacter
19	774	16.7	741	2	P95782	P95782 xanthomonas
20	585.5	12.6	711	2	Q47900	Q47900 flavobacter
21	543	11.7	723	2	Q66223	Q66223 porphyromon
22	534	11.5	723	2	Q31048	Q31048 porphyromon
23	529.5	11.4	901	3	Q96VT7	Q96vt7 aspergillus
24	510	11.0	730	2	Q93JY4	Q93jy4 prevotella
25	508.5	10.9	793	3	O14073	O14073 schizosacch
26	505.5	10.9	765	6	Q8WVG8	Q8wvg8 bos taurus
27	486	10.5	765	6	Q9N2I7	Q9n2i7 felis silve
28	477	10.3	799	5	O181I9	O181i9 caenorhabdi
29	471.5	10.1	748	13	P70092	P70092 xenopus lae
30	469.5	10.1	765	3	O14425	O14425 aspergillus
31	458	9.9	771	3	O42812	O42812 aspergillus
32	457	9.8	804	11	Q9Z218	Q9z218 mus musculu
33	456	9.8	707	16	Q9F348	Q9f348 streptomyce
34	453	9.8	755	13	Q91651	Q91651 xenopus lae
35	449.5	9.7	803	11	Q9QVT8	Q9qvt8 rattus sp.
36	444	9.6	711	4	Q9P236	Q9p236 homo sapien
37	438.5	9.4	761	11	Q8R492	Q8r492 rattus norv
38	437.5	9.4	988	5	Q9VMB4	Q9vmb4 drosophila
39	430.5	9.3	237	2	O87543	O87543 capnocytoph
40	409.5	8.8	802	5	Q9VUH1	Q9vuh1 drosophila
41	394.5	8.5	795	16	Q9PHC9	Q9phc9 xylella fas
42	368	7.9	745	5	Q9VMM2	Q9vmm2 drosophila
43	360.5	7.8	772	2	Q9AN37	Q9an37 bradyrhizob
44	357	7.7	789	2	Q93M73	Q93m73 xanthomonas
45	333.5	7.2	829	5	Q18253	Q18253 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q8WXD8	PRELIMINARY:	PRT:	863 AA.
AC	Q8WXD8:			
DT	01-MAR-2002 (TReMBLrel. 20, Created)			
DT	01-MAR-2002 (TReMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	Dipeptidyl peptidase 9.			
GN	PPP9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Olsen C., Wagtman N.;			
RT	"Identification and characterization of a novel member of the			
RT	dipeptidyl peptidase IV-related family.";			
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF452102; AAL47179.1;			
DR	InterPro; IPR002469; DPPIV_N_term.			
DR	InterPro; IPR001375; Peptidase_S9.			
DR	InterPro; IPR002410; Pro_annoPTase.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	Pfam; PF00930; DPPIV_N_term; 1.			
DR	Pfam; PF00326; Peptidase_S9; 1.			
DR	PRINTS; PR00793; PROAMNOPTASE.			
SQ	SEQUENCE 863 AA; 98263 MW; 40FE0B78E26CDED5 CRC64;			
Query Match 100.0%; Score 4646; DB 4; Length 863;				
Best Local Similarity 100.0%; Pred. No. 0;				
Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	MATGTGPTADRGDAATDDPAARFOVKRHSNDGLRSIIHGSRKYSGLIVNKAPHDFQFVQ	60	
Db	1	MATGTGPTADRGDAATDDPAARFOVKRHSNDGLRSIIHGSRKYSGLIVNKAPHDFQFVQ	60	
Qy	61	KTDESGPHSHRLYLGLMPYSGRENSLLYSEIPKKVKRKALLLLSWKQMLDHFQATPHGVG	120	


```
Qy 605 YVPPIFHFRTRSVRLYGMITYKPHALPGKKHPTVLFVGGPQVQLVNNSEFKGIKYLRL 664
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 615 YTPPEIFSEFTTGLTYLMYKPHDLQPGKKYPTVLFVGGPQVQLVNNRFGKVKYFRL 674
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 665 NTLASLVAVVVIDGRSCORGLRFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRV 724
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 675 NTLASLVAVVVIDNRSGRHLRFEKALKNQMGQVELEDQVGLQFVAEKYGDIDLSRV 734
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 725 AIHWSVGGFLSLMGLIHKPQKFAVATAGAPVTVMAYDTGTYRYMDVPENNOHGYEAG 784
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 735 GIHWSVGGFLSLMGLIHKPQKFAVATAGAPVTVMAYDTGTYRYMDVPENNOHGYEAG 794
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 785 SVALHVEKLPNEPRLILHGLFDLNDVHFFHFNFLVSLIRAGKPYQIQIYIPNERHSIRC 844
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 795 SVAMQAEKFPSEPNRLILHGLFDLNDVHFFHFNFLVSLIRAGKPYQIQIYIPNERHSIRC 854
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 845 PESGEHYEVTLHFLQEVYL 863
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 855 PESGEHYELHLLYLQENL 873
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
Q9HBM5 PRELIMINARY; PRT; 882 AA.
ID Q9HBM5 AC Q9HBM5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Dipeptidyl peptidase 8.
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
RL Eur. J. Biochem. 267:6140-6150(2000).
DR EMBL; AF221634; AAG29766.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR002469; DPPIV_N.term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N.term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 882 AA; 101421 MW; AD801C302DB4652B CRC64;
```

```
Query Match 61.88; Score 2870; DB 4; Length 882;
Best Local Similarity 61.5%; Pred. No. 1.5e-221;
Matches 517; Conservative 134; Mismatches 187; Indels 2; Gaps 2;

Qy 24 FOVOKHSDGLRSIIHSGRSKYSGLIVNKAHPDFQVOKTDSGPHSHRLYLGMYPGSGRE 83
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 FYVERYSHWSQLKLLADTRKTHGYMAKAPDFPVKNDDPGHSDRIYILAMSGENRE 94
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 84 NSLYSEIPKVRKREKALLLSWQKMLDFOATPHHGVYSREELLRERKRLGVGITSYD 143
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 NTLFYSEIPKTIINRAVLMKSLKPLDLFQATLDYGMYSREELLRERKRLGVGITSYD 154
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 144 FHSSEGLFLFQASNLFCRGRGNGFMVSPKPLEIKTQSCGPRMDPKICPADPAFTSF 203
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 YHQSGTFLFQAGSGIYHVGKGGPGGFTQQPLRPNLNVETPCNTRMDPKLCPADPDWTF 214
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 204 INNSDLWANTETGERLRTCHQGLSNVLDLDPKSGAGVATVIOEEDRFTGYWMCPTAS 263
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 IHSNDIWSINVTREERLTTVHNLANNEEDARSAGVATVFLQEEFDRTYSGIYWMCPTAS 274
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 264 WEGSEGLKTLRLYEVDSEVEIHPVSPALEERKTDTSYRPTGSKNPKIALKLPARFQ 323
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 275 TTPSGG-KILRLYEENDESEVEIHHVTSPLMLETTRADSFYRPTGTANPKVTFKMSM 333
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 324 TDSOGKIVSTOEKELQVPESSLFKPEVIARAGTRDGKYAWAMFLDRPQQLVLVLP 383
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 334 IDAGRIIDVDKELIQFEILFEGVEIARAGTWPEKGYAWSILLDRQLQIVLISP 393
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 384 ALFIPSTENESQRLASARAVPRNVQPVYVEEVTVNWINVHDIFFYPPQSGEGEDLCFLR 443
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 ELFIPVEDDVNERQRLIESVPDSVTPLIYEETTDIWINIHDIHFVFPQSH-EEIEIF 452
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 444 ANECKTGFCCHLYKVAVLKSGQYDWESEFPFSGEGEDFCPIKEETALTSGEVEVLARHGS 503
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 ASECKTGFRHLYKITSILKESKYKRSGLPAPSDFKCPIKEETAITSGEVEVLGRHGSN 512
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 504 IWNVEETKLVYFOGKDTPLLEHLYVWSYEAAGETVRLTTPGFSSHSCSMSONDFMFSHY 563
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 513 IQDEVRLVYFEGTKDSPLEHLYVWSYVNPGEVTRLTDRGYSHSCCISQHCDFISKY 572
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 564 SSVSTPPCVHYKLSGDDDDPLHKQPREWASMEAAASCPDPYVPEIFHFRTRSDVRLYG 623
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 573 SNQKNPHCVSLYKLSPPDDPTCKTEFWATILDSAGPLPDYTPPEIFSFESTTGTLYG 632
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 624 MIYKPHALQPKKHTVLFVYGGPQVQLVNNSEFKGIKYLRLNTLASGYAVVVIDGRSC 683
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 633 MLYKPHDLQPGKPYTVLFVYGGPQVQLVNNRFGKGYFRNLTLASGYVAVVVIDNRGSC 692
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 684 QRLRFEKALKNQMGQVELEDQVGLQFVAEKYGFIDLSRVAIHGWSYGGFLSLMGLIHK 743
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 693 HRGLKFEKAFYKMGQIIEIDQVGLQYLSRYDIDLRVGIHGWSTGGYLSLMALMQR 752
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 744 PQVKVATAGAPVTVMAYDTGTYRYMDVPENNOHGYEAGSVLHVEKLPNEPRLIL 803
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 753 SDIFRVAIAGAPVTLWIFYDTGTYRYMGHPDQNGYVLSGVAMQAEKFPSEPNRLILL 812
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 804 HGFLDENVHFFHFNFLVSLIRAGKPYQIQIYIPNERHSIRCPSEGEHYEVTLHFLQEVYL 863
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 HGFLDENVHFFHFNFLVSLIRAGKPYQIQIYIPNERHSIRCPSEGEHYEVTLHFLQEVYL 872
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
O75273 PRELIMINARY; PRT; 508 AA.
AC O75273;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE R26984_1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lanerudin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Carnes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Frankel M., Amico-Keller G.,
RA Cosfield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RT "Sequence analysis of a 2.5 Mb region in 19p13.3."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005594; AAC33801.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR002410; Pro_annoPase.
DR InterPro; IPR000379; Ser_estrs_site.
DR PRINTS; PR00793; PROAMNOPTASE.
ET NON_TER
SQ SEQUENCE 508 AA; 57750 MW; 2F4D6645BE2D2C89 CRC64;
```

Query Match 55.4%; Score 2572; DB 4; Length 508;
Best Local Similarity 90.8%; Pred. No. 5.6e-198;

Matches 493; Conservative 3; Mismatches 3; Indels 44; Gaps 5;

```

QY 330 IVSTQKELQPSSELPKVEYIARAGWTRDGKYAWAMELDRPOOMIQLVLLPALPIPS 389
Db 1 IVSTQKELQPSSELPKVEYIARAG-----AWAMELDRPOOMIQLVLLPALPIPS 53
QY 390 TENEEQRLASARAVPRNVQYVYVEETVNWVHDIFFYFPQSEGDELCFLRANECKT 449
Db 54 TENEEQRLASARAVPRNVQYVYVEETVNWVHDIFFYFPQSEGDELCFLRANECKT 113
QY 450 GFCHLYKVTAVLKSQGDWSEPSGDEPKCPKEIKEDIALTSGSEVILARHGSKIWNNEE 509
Db 114 GFCHLYKVTAVLKSQGDWSEPSGEG-----EQLTNA-----IWNNEE 154
QY 510 TKLVYFGTQDTPLHHLHYVVSVEAAGEIVRLTPGFSHSCSMSONDFVSHVSSVSTP 569
Db 155 TKLVYFGTQDTPLHHLHYVVSVEAAGEIVRLTPGFSHSCSMSONDFVSHVSSVSTP 214
QY 570 PCVHVYKLSGPDGDDPLHKQPRFNASMMEAAACPPDYVPPEIFHFHTRSDVRLYGMVYKPH 629
Db 215 PCVHVYKLSGPDGDDPLHKQPRFNASMMEAA-----KIFHFHTRSDVRLYGMVYKPH 265
QY 630 ALQPKKKHPTVLFYVGGPQVLVNNSEFKIKYLRNLTLASLGAVVVDGRGSCQGLRF 689
Db 266 ALQPKKKHPTVLFYVGGPQVLVNNSEFKIKYLRNLTLASLGAVVVDGRGSCQGLRF 325
QY 690 EGALKNMOMGOVEIDQVEGLQFVAEKYGFIDLRSVATHGWSYGGFLSLMGLIHKPQVFKV 749
Db 326 EGALKNMOMGOVEIDQVEGLQFVAEKYGFIDLRSVATHGWSYGGFLSLMGLIHKPQVFKV 385
QY 750 AIAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNPNRLLILHGFLE 809
Db 386 AIAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAGSVLHVEKLPNPNRLLILHGFLE 445
QY 810 NVHFFHTNFLVSLIRACKPYQL-----QIYPNERHSIRCPESGEHYEVTLLHFLQ 860
Db 446 NVHFFHTNFLVSLIRACKPYQLQVALPPVSPQIYPNERHSIRCPESGEHYEVTLLHFLQ 505
QY 861 EYL 863
Db 506 EYL 508

```

```

RESULT 5
Q9BVR3 ID Q9BVR3 PRELIMINARY; PRT; 439 AA.
AC Q9BVR3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 49.9 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC000970; AAH00970.1; -.
DR MEROPS; S09.019; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser.estrs_site.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 439 AA; 49926 MW; A18BBA9E12092BAF CRC64;

Query Match 51.3%; Score 2383; DB 4; Length 439;
Best Local Similarity 100.0%; Pred. NO. 6.9e-183;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 425 DIFYPPQSGEGDELFLRANECKTGFCCHLYKYTAVLKSQGDWSEPFSGEDEFKCPK 484
Db 1 DIFYPPQSGEGDELFLRANECKTGFCCHLYKYTAVLKSQGDWSEPFSGEDEFKCPK 60
QY 485 EETALTSGEWEVLARHGSKIWNNEETKLVIYFGTQKTPLHHLHYVVSVEAAGEIVRLTTP 544
Db 61 EETALTSGEWEVLARHGSKIWNNEETKLVIYFGTQKTPLHHLHYVVSVEAAGEIVRLTTP 120
QY 545 GFSHSCSMSONDFVSHVSSVSTPCVHVYKLSGPDGDDPLHKQPRFNASMMEAAACPPD 604
Db 121 GFSHSCSMSONDFVSHVSSVSTPCVHVYKLSGPDGDDPLHKQPRFNASMMEAAACPPD 180
QY 605 YVPEIFHFHTRSDVRLYGMVYKPHALQPKKHPTVLFVYGGPQVLVNNSEFKIKYLR 664
Db 181 YVPEIFHFHTRSDVRLYGMVYKPHALQPKKHPTVLFVYGGPQVLVNNSEFKIKYLR 240
QY 665 NTLASLGAVVVDGRGSCQGLRFEGALKNMOMGOVEIDQVEGLQFVAEKYGFIDLRSV 724
Db 241 NTLASLGAVVVDGRGSCQGLRFEGALKNMOMGOVEIDQVEGLQFVAEKYGFIDLRSV 300
QY 725 AIHGWSYGGFLSLMGLIHKPQVFKVATAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAG 784
Db 301 AIHGWSYGGFLSLMGLIHKPQVFKVATAGAPVTVMAYDGTGYTERYMDVPENNQHGYEAG 360
QY 785 SVALHVEKLPNPNRLLILHGFLENDVNHFFHTNFLVSLIRACKPYQLQIYPNERHSIRC 844
Db 361 SVALHVEKLPNPNRLLILHGFLENDVNHFFHTNFLVSLIRACKPYQLQIYPNERHSIRC 420
QY 845 PESGEHYEVTLLHFLQBYL 863
Db 421 PESGEHYEVTLLHFLQBYL 439

```

```

RESULT 6
Q96JX1 ID Q96JX1 PRELIMINARY; PRT; 632 AA.
AC Q96JX1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ14920 fis, clone PLACE1007416, weakly similar to dipeptidyl
DE Peptidase IV (EC 3.4.14.5).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027826; BAB55395.1; -.
DR MEROPS; S09.018; -.
SQ SEQUENCE 632 AA; 72639 MW; 9BDF598B06985AA4 CRC64;

Query Match 46.0%; Score 2137.5; DB 4; Length 632;
Best Local Similarity 58.5%; Pred. NO. 6.3e-163;
Matches 395; Conservative 93; Mismatches 134; Indels 53; Gaps 3;

```

```

QY 189 MDPKICPADPAFFSFINNSDLWANITGERRLTFC HQGLSNVLDPPKSGAGVATFVIOE 248
Db 1 MDPKICPADPDWIAFTHSNDIWSNIVTREERRLT VYVHNLANNEDARSAGVATFVLQE 60
QY 249 EFDRFYGGWCPATSGESGLTKTLRLIYEVESEVEVHVPSPALEERKTSYRPT 308
Db 61 EFDYSGYWCPCAEETTPSGG-KILRLIYEENDESEVEIIVHTSPMLETRADSFYRPT 119

```



```
Db 121 YAVVVVDGRGSCQGRUFEGALKNQGVIEDQVEGLOFVAEKYGFIDLSRVAIHGSY 180
QY 732 GGFSLMGLLHKQVQKVAITAGAPVTVMAYDTGYTRYMDVPPENNOHGYEAGSVLHVE 791
Db 181 GGFSLMGLLHKQVQKVAITAGAPVTVMAYDTGYTRYMDVPPENNOHGYEAGSVLHVE 240
QY 792 KLPNEPNRLILHGFGLDENVHFHTFLVSOLIRAGKPYQLOIYTPNERHSIRCPESGEHY 851
Db 241 KLPNEPNRLILHGFGLDENVHFHTFLVSOLIRAGKPYQLOIYTPNERHSIRCPESGEHY 300
QY 852 EVTLHLFLQBYL 863
Db 301 EVTLHLFLQBYL 312

RESULT 9
Q9VC20 PRELIMINARY; PRT: 1042 AA.
AC Q9VC20;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG3744 protein.
GN CG3744.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aquayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.R., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Poilard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DB EMBL; AE003749; AAF56357.1;
```

```
DR MEROPS; S09.016; -.
DR FlyBase; FBgn0039240; CG3744.
DR InterPro; IPR002469; DPPIV_N_term.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs_site.
DR Pfam; PF00930; DPPIV_N_term; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
SQ SEQUENCE 1042 AA; 116706 MW; 5B915D5C365DF937 CRC64;

Query Match 35.1%; Score 1628.5; DB 5; Length 1042;
Best Local Similarity 39.1%; Pred. No. 1.1e-121;
Matches 370; Conservative 143; Mismatches 315; Indels 119; Gaps 20;

QY 3 TTGTP---TADRGAAA-----TDDPA-----ARFQVQKHSWDGLRSIIHGRKYSGL 47
Db 124 TSGTPPHGLDVDEGDDEDDPVDNDGHIAAPTNPKSWAENKQVQVIRKKMCN 183
QY 48 IVNKAPHDFQVOKTDESGPHSHRLYYLGMYPGSRNSLLYSIIPKVKREALL-----102
Db 184 LSSMVPNTNQFRHLSGDA-----RCYFLGTPPOSWETTLFADINLTQSEQQLLVQRL 239
QY 103 -----LSWKOMLDH-FOATPHHG-----VY 121
Db 240 GIASDEWSPTMNAGSPTSSGHOPAFLENSLPRPLPWSLLOQPIQSSGGSGSGSASPY 299
QY 122 SREEELLRRKRLGVFGITSYDFHSESGFLFOASNSLHFCRDGKGNFMVSPMKPLEIK 181
Db 300 AREYQLLOERKRLSTWGIISYELHKPSGLVPCFNDLYQCLDTGYNSGL---LFTQLR 356
QY 182 TQSGRPMDFKICPADPAFFSFNNSDLMVANIETGEERRLTFCHOGLSNVLDDPKSAGV 241
Db 357 TCPQWTALDPQICPNQSDMIAYISDCDLFTVHTLSGHEKRLTYTSGRHSYVDLALSAGV 416
QY 242 APTVIOEEEDRFTGYWMCPTASWEGSEGLKTLRLIYEEDESEVEYIHPSPALEERKTD 301
Db 417 PSYVMOEEFSRYGFWQMP-----HSNDGI--YRIVVEEVESEVSYTTPSSAMHGRVD 470
QY 302 SYRYPRTSGKNPKIALKLAEFQDSQGIKIVSTQEKELVQFPFSLFPKVEYIARAGTRDG 361
Db 471 EYRFPRTSPNAKSKLKVQLVNEALQVSEITAKDLPYSLLAVFSWLEIVYRGTWTPDA 530
QY 362 KYAWMFLDRPQWLQVLVLPALFIPSTENEEQRLASA-----RAYPRNVQP-VYVY 413
Db 531 KYVWVGGLDRKQKQDLVILIPDNFCESYSSQVSTPTDSIGDHSWRSLSRTITPLQVY 590
QY 414 EEVNTVINVDHIFYPPOSEGEDELCELRANCKTGCHLYKVTA-VLKSQGYDWSEPF 472
Db 591 TERSDSWINVHDMHLFDLT--ETSVTFWASB-ETGFRHLVLTASLLLSQANGQDPG 647
QY 473 SPGE-----DEFKCPIKEIALTSGEWEVLARHGSKIWNNEETKLIVYFQGTDPLEH 525
Db 648 SVGAQPSFVDSLALQPRILNKVALTSGEVELARN---LWVDKANKLVYFVGURDTPLEK 704
QY 526 HLYVVSYEAGEIVRLTTGFSHSCSMSONDFVSHYSVSTPFCVHYVKKLSPGDDP- 584
Db 705 HLYVVSLEPERHILRLTEFGYSYLVFEFDQCKLMLLVLYCNIQRLPSCCKVMRNVQTSNGG 764
QY 585 -----LHKQPRFWASMEAAACPDYVPPPIEFHFTFSDVRLYCYMIYKHALQP 633
Db 765 VNGIQISLVGYLH-----EGGKPEPYC-POIFSPQLPSGDIVYAWFKPINFEL 813
QY 634 GRKHPTFLVYGGPQVOLVNNFSGIKYLRANTLASLYAVVYIDGRSCQRLREGAL 693
Db 814 GVKYPTVLNVYGGPEVQVTNNTFKGKHLRMHMLAAQGCVCICIDSRGSRHRKREFSHI 873
QY 694 KNOMGOVEITDQVEGLOFVAEKYGFIDLSRVAIHGSYGGFSLMGLIHKPQVFKVAIAG 753
Db 874 RGRMGVELTDTQDALRSLSDQLGYIDMDRVATHGWSYGYLSGLMGLVQYPKTFKVAIAG 933
QY 754 APVTVMAYDTGYTRYMDVPPENNOHGYEAGSVLHVEKLPNEPNRLILHGFGLDENVHF 813
Db 934 APVTVMAYDTGYTRYMDVPPENNOHGYEAGSVLHVEKLPNEPNRLILHGFGLDENVHF 993
```

Matches	370;	Conservative	143;	Mismatches	315;	Indels	119;	Gaps
QY	3	TTCTP----	TADRGAAA-----	TDDPA-----	ARFQVQKHSMDGLRSIIHSGRKYSGL	47		
Db	184	TSCTPHGLDVGDEDCDEDEDDPVDNDGHIAATPNKSWAENKQVQVEIRKKWCN	243					
QY	48	IVNKAHDFQVQKTDSPGHSHRLYYLGMYPYSGRENSLLYSEIPKPKYKREKALL-----	102					
Db	244	LSSMVPTNVQFRHLSGDA-----RCYFLGTTPQSWETTLFADINLTQSEQQLLVQRL	299					
QY	103	-----	LSWQKMLDH-FQATPHHG-----	VY	121			
Db	300	GIASDEWSPMTNAGSPSTSSGHQPAFLFNSLPRPLPWSPLQOQPQSSGGSGGSSASPY	359					
QY	122	SREELLREKRLGVGITSYDFHSESGLFLFOASNSLFLHCRDGGKNGFMVSPMKPLK	181					
Db	360	AREYQLLOERKRLSTWGTITSYELHFKPSGKLVPFCFNDLYQCLDTGYNLGL-----	416					
QY	182	TQCSGPRMDPKICPADPAFFSFINNSDLWVANIETGEERLTFCHQGLSNVLDDPKSAGV	241					
Db	417	TCQWTALDPQICPQNSDMIAYISDCDLFTVHTLSGHEKRLTYTSTGRHSYVDALSGV	476					
QY	242	ATFVIOEEDRFTGYWMCPTASWEGSEGLKTLRIIYEEVDESEVEVIVHPSPALEKTD	301					
Db	477	PSVMOEEFSRYQGFWMQP-----HNDGI--YRIYEEVDESEVYTFPSTAMHGRVD	530					
QY	302	SYRPTGSKNPKIALKAEFOTDSOGKIVSTQEKELQVPSFSLPKPKVEYIARAGWTRDG	361					
Db	531	EYFPRTGSPNAKSLKLVQFVNLALQVSEAIKDLPSYLLAVSWLEYIVRVGWTPTDA	590					
QY	362	KYAWMFLDRPOOWLOLVLLPALFIPSTENEEQRLASA-----RAVRPNVQV-YVYV	413					
Db	591	KYVWQGLDRKQRLDVLILPDNFCESYSQSVSTPDSIGDHSWRSLSYRTITPLQVIY	650					
QY	414	EVTNVWVNDHIFYPFPOSEGEDELFLRANECKTGCHLYKYVA-VLKSQGYDWSFPF	472					
Db	651	TERSDSMNVDHMLHFLDLT--ETSVTFLWASE-ETGFRHLYLVTASLLSOANGQPDGP	707					
QY	473	SPGE-----DEFKCPKEEIALTSGEWEVLARHGSKIWNNEETKLVYFOGKTDPLEH	525					
Db	708	SVGAQPSFVDLSALQPLINKVALTSGEWEVLARN---LWVDKANKLYFVGLRDTPLEK	764					
QY	526	HLVYSYEAGEIVRITTPGFSHSCSMQNFDMVSHVSSYSTPCVHVYKLSGDDDP--	584					
Db	765	HLVWSLERPEHRLITTEPGYSYLVFEDDOCKMLLVYCNQIRLPSCKVMRVNQTCSNGG	824					
QY	585	-----LHKQPRFWASMMEAASCPDYVPPEIFHFHTRSVRLYGLMYKPHALQP	633					
Db	825	VNGIQISLVGLH-----EGGKPEPYC-QPISQPLSPGSDIVYAMVFKPHNFEL	873					
QY	634	GKKHPTVLFYVGQPVQVNNNSFKIKYLRNLNTLASLGYAVVVIDGRSGCORGLFEGAL	693					
Db	874	GVKYPVLVNYGPEVQTVNNTFKGKHQLRMHMLAAQGYCVCIDSRSGRHRGKFESHI	933					
QY	694	KNQMGQVEIEDQVEGLQFVAEKYGFIDLSRVAIGHWSYGGFSLSLMGLIHKPQVFKVAIAG	753					
Db	934	RGMRGQVELTDQVDAURLSLDQGLYIDMRVAIGHWSYGGYLSLGLVQYKPKIFKVAIAG	993					
QY	754	APVTVMAYDVTGTYERYMDVPPNNQHYEAGSVALHVEKLPEPNRLLILHGLFDENVHF	813					
Db	994	APVTNMEYDVTGTYERYMDPQNNNEAGYSAGSVLEYSVFPFEEDKRLILLIHLIDENVHF	1053					
QY	814	FHTNFIIVSOLIRAGKPYQLOIYVNERHSIRCPESGEHYEVTLLHFLQ	860					
Db	1054	CHTSRLISALNANKANKPYEVLHFPFEEHRSURNLESNNKNETKLLSLFQ	1100					
RESULT	11							
Q9HBM3								
ID	Q9HBM3							
AC	Q9HBM3							
DT	01-MAR-2001 (TREMBLrel. 16, Created)							
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)							

```
DT 01-DEC-2001 (TreeBLrel. 19, Last annotation update)
DE dipeptidyl peptidase 8 (fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DR EMBL: AF221636; AAG29768.1; -.
DR MEROPS: S09.018; -.
FT NON_PER 1
SQ SEQUENCE 465 AA; 53197 MW; 22939EC0A4FE29CE CRC64;

Query Match 32.2%; Score 1494.5; DB 4; Length 465;
Best Local Similarity 50.5%; Pred. No. 1.8e-111;
Matches 281; Conservative 69; Mismatches 105; Indels 101; Gaps 2;

QY 308 TGSKNPKIALKAEFQDSQKIVSTQEKELVQPFSSLPFKVEYIARAGWTRDGYAWAM 367
DB 1 TGTANPKVTFKMEIMIDABRIIDVIDKELIQPFELFEGVEYIARAGWTPGKYAWSI 60

QY 368 FLDRPQWLQVLVLPALFIPSTENEOQLASARAVPRNPVQVYVEETVNWVHDF 427
DB 61 LLRSQTRQLVLSPELFIPVDDVMERQLIESVPDSVTPLIYEETDWINIHDF 120

QY 428 YPPQSEGEDELFLANECKTCFCHLYKVTAVLKSQGDWSEPFSPGDEKCPKEEI 487
DB 121 HVPQSH-EEIEFIFASECKTGRHLYKITSILKESKYRSGGLPAPSDFCPKREEI 179

QY 488 ALTSGEWEVLARGSKIWNEETKLVYFQGTDPLEHLYVYVYEAAGIVRLTTPGES 547
DB 180 AITSGEWEVLGRHGSNIQDVRRLVFECKTDSPLEHLYVYVYVNPGEVRLTRDGY 239

QY 548 HSCSMQNFDMFVSHVSSVSTPCVHVYKLGGDDPLHKQPRFASMEAAACPDYVP 607
DB 240 HSCCIHQHCDFFISKYSNQKNPCVSLYKLSSPDPTCKTFEATILDSAGLPDYTP 299

QY 608 PELFHFHTRSDVRLYGMVYKPHALQPKHPTVLYVYGGPOVLVNSFKGIKYLRLNTL 667
DB 300 PELFSFESTGTGLYGLMYKPHDLQPKGYPTVLYVGGPQ----- 340

QY 668 ASLGYAVVVIDGRSGCQGLRFBGALKNQMGVEIEDQVEGLQFVAEKYGFIDLSRVAIH 727
DB 341 ----- 340

QY 728 GWSYGGFSLMGLHKKPQVKVAIAGAPVTVMAYGTGTYTERMDVNPENNOHQYAGSVA 787
DB 341 -----VAIAGAPVTWIFDTGTYTERMGHPDQNEQCYLGSVA 379

QY 788 LHVEKLPNEPRLILHGLFDENVHFFHNFVLSQLIRAGKPYQLQIYPNERHSIRCPES 847
DB 380 MQAEKFPENRLLHGLFDENVHFAHTSILLSFLVRAGKPYDLQIYPOERHSIRVPES 439

QY 848 GEHYEYTLHLFLQEYL 863
DB 440 GEHYELHLLHYLOENL 455

RESULT 12
Q9NMF4
ID Q9NMF4 PRELIMINARY; PRT; 469 AA.
AC Q9NMF4;
DT 01-OCT-2000 (TreeBLrel. 15, Created)
DT 01-OCT-2000 (TreeBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TreeBLrel. 19, Last annotation update)
```

```
DE CDNA FLJ20283 fis, clone HEP04088.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DR EMBL: AK000290; BAA91059.1; -.
DR MEROPS: S09.018; -.
SQ SEQUENCE 469 AA; 54367 MW; 088EED0B1E46C11F CRC64;

Query Match 29.6%; Score 1376; DB 4; Length 469;
Best Local Similarity 56.1%; Pred. No. 6.1e-102;
Matches 245; Conservative 85; Mismatches 105; Indels 2; Gaps 2;

QY 24 FOVKHSDGLRSIIHSGSRKYLIVNKAPHDFQFQVQKTDSEGHRLYYLGMPTGSR 83
DB 35 FYVERYSWQKLLADTRKYHGYMMAKAPDFMFVKNRNDPDPGPHSDRIYYLAMSGENR 94

QY 84 NSLLYSEIPKKVKKEALLLSWKQMLDHFQATPHGHVYSREELLRKRKLVGFGITSYD 143
DB 95 NTLFYSEIPKTNRAAVLMSWKPLDLDFQATLDYGMYSREELLRKRKRGTVGSIASVD 154

QY 144 FHSSEGLFLFQASNSLFCRDCGKNGFMVSPMKPLIKTCQCSGRMDPKICPADPAFFSF 203
DB 155 YHOGSGTFLFQAGSGLYHVKGDPGQFTQOPLRPNLVETSCPNIRMDPKLCPADPDWIAF 214

QY 204 INNSDLWANIETGERRLTFCRHSNVLDDPKSAGVATFVIOEEDFDRFTGYWMCPTAS 263
DB 215 IHSNDWISNIVTREERLTYVHNELANEEDARSAGVATFVIOEEDFDRYSGVWCPKAE 274

QY 264 WEGSEGLTLRLIYEEVDESEVEVHVSPALAEERKTSYRPTGSKNPKIALKLAEFQ 323
DB 275 TTPSGG-KILRLIYEEVDESEVEIHWTPMLETRRADSFYRPTGTANPKVTFKMEIM 333

QY 324 TDSOGKIVSTQEKELVQPFSSLPFKVEYIARAGWTRDGYAWAMFLDRPQWLQVLVLP 383
DB 334 IDAEGRIIDVIDKELIQPFELFEGVEYIARAGWTPGKYAWSILLSRQTRQLVLSIP 393

QY 384 ALFIPSTENEOQLASARAVPRNPVQVYVEETVNWVHDFIYFPQSEGEDELCLFLR 443
DB 394 ELFIPVEDDVMERQLIESVPDSVTPLIYEETDWINIHDFHVPQSH-EEIEFIF 452

QY 444 ANECKTGFCHLYKVTAV 460
DB 453 ASECKTGFRLHYKITSI 469

RESULT 13
Q9HBM2
ID Q9HBM2 PRELIMINARY; PRT; 360 AA.
AC Q9HBM2;
DT 01-WAR-2001 (TreeBLrel. 16, Created)
DT 01-WAR-2001 (TreeBLrel. 16, Last sequence update)
DT 01-WAR-2002 (TreeBLrel. 20, Last annotation update)
DE Dipeptidyl peptidase 8 (fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
RA Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
RT dipeptidyl peptidase (DPP) IV homolog, DPP8."
DT 01-DEC-2001 (TreeBLrel. 19, Last annotation update)
```

DR EMBL; AF221635; AAG29767.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
FT NON_TER 1
SQ SEQUENCE 310 AA; 35396 MW; 8E87C34026D9C7AC CRC64;

Query Match 23.5%; Score 1093.5; DB 4; Length 310;
Best Local Similarity 57.3%; Pred. No. 1.6e-79;
Matches 200; Conservative 42; Mismatches 58; Indels 49; Gaps 1;

QY 515 FQGTDTPTLEHLLHYVVSAAAGEIVRLTTPGFSSHSCSMSONDFMVFVSHYSYSTPPCVHV 574
Db 1 FEETKDSPLEHLLHYVVSYNPGVEVTLTRDGRGSHSCCSQHCHDFISYSNQKNPHCVSL 60
QY 575 YKLGGDDPLHKOPRFWASMMEAASCPDPYVPPEIFHFHTSDVRLYGMIXKPHALOPG 634
Db 61 YKLSSDEDDPTCKTEFNATILDSAGLPDYTPPELFSESTGTFLYGLMYLKPHDLQPG 120
QY 635 KKHPTVLVYVGQPQLVNNSFKIKYLRLNTLASLGVAVVVIDRGSGQRGRLEFEGALK 694
Db 121 KKYPVTLVFYGGPQ----- 134
QY 695 NMQGVEIEDQVEGLQFVAKYGFIDLSRVATHGSYGGLSLMGLIHKPQVKYAIGA 754
Db 135 ---GQIEIDDQVEGLQYLAASYDFIDLDRVIGHWSYGGLYSLMALMQRSDFIRVAIGA 191
QY 755 PVTVMAYDTGYERYMDVNPENHQGVAGSVALKVEKLPNEPNRLLILHGFLDENVHFF 814
Db 192 PVTLMIFYDTGYTERYMGHPDQNEQQYYIGSVAMAQEPSPNNKLLLHLGFLDENVHFA 251
QY 815 HTNFLYSOLIRAGKPQLOIYIPNERHSIRCPSGEHVEYVTLHFLQEYL 863
Db 252 HTSILLSLFLVRAGKPYDIQIYPOERHSIRVPESGEHYELHLLHYLQENL 300

RESULT 15
Q965K3 PRELIMINARY; PRT; 927 AA.

ID Q965K3 AC Q965K3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-MAR-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein K02F2.1b.
GN K02F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoidea; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Maggi L., Goela D.;
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL The sequence of C. elegans cosmid K02F2.";
Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AF043699; AKK84627.1; -.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrns_site.
DR Pfam; PF00326; Peptidase_S9; 1.
DR

R. J. Biochem. 267:6140-6150(2000).
EMBL; AF221637; AAG29769.1; -.
DR MEROPS; S09.018; -.
DR InterPro; IPR001375; Peptidase_S9.
DR Pfam; PF00326; Peptidase_S9; 1.
FT NON_TER 1
SQ SEQUENCE 360 AA; 41070 MW; CF81COBBG1423E22 CRC64;

Query Match 26.3%; Score 1220.5; DB 4; Length 360;
Best Local Similarity 44.2%; Pred. No. 1.3e-89;
Matches 236; Conservative 54; Mismatches 69; Indels 175; Gaps 2;

QY 234 DPKSAGVATFVIQEEFRFTGYWCPTASWGSEGKLTRLRYEVDSEVEIHVSP 293
Db 2 EDARSAGVATEVLQEEFRDTSYWCPCAEATPSSG-KILRLIYEENDESEVEIIHVTSP 60
QY 294 ALBERTDSYRPTGSKNPKALKLAEBOTDSQKIVSTQBELVQPFSSLFPKVEXIA 353
Db 61 MLETNRADSRFPKTGTANPKVTFKMSIMDAEGR-- 98
QY 354 RAGWTDRGKYAMFLDRPQWLQVLVLPALFIPSTENEORLASARAVPNQPVVY 413
Db 99 ----- 98
QY 414 EEVTNWINVHDIFYPPPOSEGEDELCLFLANECKTGCHLYKVTAVLKSQGYDWSEPS 473
Db 99 ----- 98
QY 474 PGDEFKCIKEIALTSGEVENLARHSGKIWNDETKLIVFOGTQTPTLEHLLHYVSYE 533
Db 99 -----VDEVRLVYFEGTKDSPLEHLLHYVSYV 126
QY 534 AAGEIVRLTTPGFSSHSCSMQNDFMVSHYSVSTPPCVHVVYKLSGPDDPLHKOPFWA 593
Db 127 NPGEVTRLTRDGSHSCCSIQCHDFISYSNQKNPHCVSLKLSPPEDDPTCKTKFEWA 186
QY 594 SMMEAASCPDVPYPPPIFHFTRSVRLYGMIXKPHALQPKKKHPTVLVFGGQVQLVN 653
Db 187 TILDASGPLDPVTPPEIFSFESTGTFTGLYMLYKPHDLQPKKYPVTVFIYGGQVQLVN 246
QY 654 NSFKGKYLRLNTLASLGVAVVVIDRGSGQRGRLEFEGALKNQMGVEIQVEGLQVFA 713
Db 247 NREKGVKYFRLNLASLGVAVVVIDNRGSGRGLFEKAFYKMGQIIDBQVEGLQVLA 306
QY 714 EKYGFDLSRVATHGSYGGLSLMGLIHKPOVKVAIAGAPVTVMAYDTGYT 767
Db 307 SRDYFDLDVRGIGHWSYGGLYSLMALMQRSDFIRVAIAGAPVTLWIIFYDTGYT 360

RESULT 14
Q9HBM4 PRELIMINARY; PRT; 310 AA.

ID Q9HBM4 AC Q9HBM4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Diethylglycyl peptidase 8 (Fragment).
GN DPP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=20467194; PubMed=11012666;
RA Abbott C.A., Yu D.M.T., Woollatt E., Sutherland G.R., McCaughan G.W.,
Gorrell M.D.;
RT "Cloning, expression and chromosomal localization of a novel human
diethylglycyl peptidase (DPP) IV homolog, DPP8.";
RL Eur. J. Biochem. 267:6140-6150(2000).

```
SQ SEQUENCE 927 AA; 105365 MW; 6907C6AAEF829D0F CRC64;

Query Match      20.3%; Score 943.5; DB 5; Length 927;
Best Local Similarity 29.4%; Pred. No. 9.6e-67;
Matches 263; Conservative 146; Mismatches 316; Indels 169; Gaps 34;

Qy 72 LYILGMPYGSRENLLYSEIPKVY-----RKEALLLSWKQMLDHFQATPHHGVYSR 123
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 82 MYAIISSVPGTNTQISFTIPLELVEKAQVADRKFELKLSAGYNVDYSYIRKTPPSAEFTL 141
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 124 EELLREKRKRGVGTISYDFHSEGLFLFOASNSLFH----- 161
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 142 QCE--RORSQV-VTGISDIYEI--RNGKMLMAGDQLFRYNPLNEALAAPIAVPDDQSST 196
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 162 -----CRDGGKNGFMVSPMKPLEIK---TQCSPRMDP-----KI 193
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 197 EPMDISGTSITGKGCSEAPQOSTVPPVTRIPRIKPTTSTKPTATAPTNNFVSSAKV 256
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 194 CPADPAFFSFINNSDLWANIETGEERRLTFCHOGLSNVLDLDPKSAGVATFVIOEEDRF 253
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 257 CPADSSLLAYVLNKQVYI-----EKNKGIHRTSSN--SKHITNGVPSYIVOEELERF 307
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 254 TGYWMCPTASWEGSEGLTLRLIYEEVDESEVEV----IHVPSPALEERKTDYSYRPTG 309
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 308 EGIWW-----SES-KT-RLLYEHVNEEKVAESQFGVNGDPEVA-----PMKYPRAG 351
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 310 SKNPKIALKLAEFOTDSQGIKIVSQEKELVQPFSSLPKVEYIARAGWTRDQKYAWAMFL 369
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 352 TKNAYSTLRVILE--NGRAYDPLKDEV--IYKHCPEFYETIRAGFFSDGTTVWVQVM 406
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 370 DRPOOWLQVLPLPALPISPE-----NEEORLASARAVPR--- 405
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 407 SRDAQCSLLIPIYDFLLPEELGGSIKEDNLQLSTDLNMGVMDKSHETMEKPPRGKL 466
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 406 --NVQPVVYVEEVTNVINVDHIFYPFQSEGEDELQ-FLRANECKTGFCCHLYKVTAVLK 462
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 467 RGTQV---IHKARNDYWINTHNAIYPLKIIDEEHPMYEFYICLEKPNKNGSC-LALISAE 522
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 463 SOGYDNSEPFSPGDEPKCTKEIALTSGEWEVLARHGSKIWNNEETKLVIYFOGTDTP 522
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 523 QNGY-----CRHTEKLLMAENFSINKSMG--IVVDEVRELVIYVANESH 566
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 523 LEHLIYVYVEAGEIYVRLTTPGFSSHCSMSQN-----FDM----FVSHYSSVSTPPCVH 573
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 567 TEWNI-CVSHYRTGQHAQLTESGI--CFKSERANGKLALDLHDHGFACYMTSVGSPAECR 622
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 574 VYKLSGDDDLPHKQPRFWASMEASCP--PD--VYVPEIFHPT-RSDVRLYGMIXP 628
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 623 FYSPRWKENEVL--PSTVYAANITVSGHPGQPDJHFDSPEMIEFQSKKTGLMHYAMILRP 680
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 629 HALQPGKKHPTLVFVYGGPOVLVNNFSGIKYLRNLTLASLGYAVVYVVDGRCSCQRLR 688
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 681 SNEDPYKKYPVFHYVYGGPGIIVHNDFSWQIYR---FCRLGYVVVYFIDNRGSAHRGIE 737
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 689 FEGALKNQMGQVETEDQVEGLQFVAEXY-GFIDLSRVAIHGWSYGGFSLMGLIHKPOVF 747
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 738 FERHIHKMGTVETEDQVEGLQMLAERTGGFMDMSRVVHGWYSGYGNALQMIKHPNIY 797
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 748 KVATAGAPVTVMWADYDGYTERYMDVPENNQHGYEAGSVALHVEKLPNEPRLILHGL 807
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 798 RAAJAGGAVSDWRLDYFAYTERYNGYP-LBEEVYVGASSITGLVEKLPDEPNRLMLVHGLM 856
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 808 DENVHFHTNPLVSQLIRAGKPYOLOIYPNERHSIRCPESGEHYEVTLLHFLOE 861
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 857 DENVHFAHLTHLVDCEIKKGKWHELVIFPNERHGVNRNDASIYLDARMYFAOQ 910
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
```

Search completed: December 12, 2002, 12:05:51
Job time : 35 secs

```
QY 329 LysIleValSerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPheProLys 348
DB 842 TCAGTCAACCAATGCAACATTCATACAAATCACTGCTGCTTCTATGTATGATGGGAT 901
QY 349 ValGluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAlaMetPhe 368
DB 902 ---CACTACTTGTGTGATGTCAGTGGCAACACAAAGAATAATTCT----- 946
QY 369 LeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAlaLeuPheIlePro 388
DB 947 -----TTCAGTGGCTC----- 958
QY 389 SerThrGluAsnGluGlnArgLeuAlaSerAlaArgAlaValProArgAsnValGln 408
DB 959 -----AGGAGGATTCAG 970
QY 409 ProTyrValVal-----TyrGluGluValThrAsnValTrpIleAsnVal 423
DB 971 AACTATTCGGTCATGGATATTGTGACTATGATGAATCCAGTGGGAAGATGG----- 1021
QY 424 HisAspIlePheTyrProPheProGlnSerGluGlyGluAspGluCysPheLeuArg 443
DB 1021 ----- 1021
QY 444 AlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSer 463
DB 1022 -----AACTGCTTAGTGGCAGCGCAACACATGAATGAGTACTACT----- 1063
QY 464 GlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGlu----- 478
DB 1064 -----GGCTGGTGGGAAGATTTAGGCCTTCAGAACCTCATTTACCCCTTGATGGT 1114
QY 479 -----PheLys 480
DB 1115 AATAGCTTCTCAAGATCATCAGCAATGAGAGGTTACAGACACATTTGCTATTTCCAA 1174
QY 481 CysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArgHis 500
DB 1175 ATAGATAAAAGAGATGCACATTTATTACAAAGGACCTGGGAAGTCATCGGGATAGAA 1234
QY 501 GlySerLysIleTrpValAsnGluLuthrLysLeuValTyrPheGlnGlyThrLysasp 520
DB 1235 GCTCTA-----ACCAAGTGTAT 1249
QY 521 ThrProLeuGluHisHisLeuTyrValValSerTyrGluAlaAlaGly----- 536
DB 1250 -----TATCTATACTACATTAAGTATGATGATATATAAAGGAATGCCAGGAGGA 1294
QY 537 ---GluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsn 555
DB 1295 AGGAATCTTTATAAAATCCAACTTAGTACTATACAAAGTGACATGCTCAGTTGTGAG 1354
QY 556 PheAspMetPheValSerHisTyrSerValSerThrProProCysValHisValTyr 575
DB 1355 CTGAATCCGGAAGGTGTGACTACTATTCTGTGCTCATTCAGTAAAGCGCAAGTATTAT 1414
QY 576 LysLeu-----SerGlyProAspAspProLeuHis----- 586
DB 1415 CAGCTGAGATGTCGGGCTGGTCTGCTCTCTACTCTACAGCAGCGCTGAATGA 1473
QY 587 -----LysGlnPro---ArgPheTrpAlaSerMetMetGluAlaAlaSerCysPro 602
DB 1474 TAAAGGCTGAGAGTCTCGGAAGACAATTCAGCTTTGGATAAAATGCTGCAGAAATGT--- 1530
QY 603 ProAspTyrValProProGluIlePheHisPheHisThrArgSerAspValArgLeu-Ty 622
DB 1531 CCAGATGCCCTCCAAAACAGTACTTATTATTTTGAATGAACAAAATTT---TTGCTA 1587
QY 622 rGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPh 642
DB 1598 TCAGATGATCTGCTCCTCTCAT---TTTGATAAATCCAGAAATATCTCTACTATTAGA 1644
QY 642 eValTyrGlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrIle 662
```

```
DB 1645 TGTGTATGCAGGCCCATGTAGTCAAAAAGCAGACACTGTCTTTC----- 1687
QY 662 uArgLeuAsn-----ThrLeuAlaSerLeuGlyTyrAlaValValVal---IleAs 678
DB 1688 -AGACTGAAGTGGGCCACTTACCTTGCAGAGCAGACAAAACATATAGTAGTCTTGA 1746
QY 678 pGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetG1 698
DB 1747 TGGCAGAGAACTGTTACCAAGGAGATGAATCATGCGATCAATCAACAGAAAGCTGG 1806
QY 698 yGlnValGluIleGluAspGlnValGluGlyLeu---GlnPheValAlaGluLysTyrG1 717
DB 1807 AACATTGAAGTTGAAGATCAAAATTCAGAGCAGCCAGACAATTT-----TCAAAAATGG 1860
QY 717 yPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLe 737
DB 1861 ATTTGTGGACAACAACAAATTTGGGGCTGGTTCATATGGAGGTACGTAAACCTC 1920
QY 737 uMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValTh 757
DB 1921 AATGCTCTGGGATCAGGAAGTGGCGTGTCAAGTGTGAATAGCGTGGCGCTGTATC 1980
QY 757 rValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVal-----ProG1 775
DB 1981 CCGGTGGGAGTACTATGACTAGTGTACAGACAGCGTTACATGGGTCTCCCAACTCCAGA 2040
QY 775 uAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAs 795
DB 2041 AGACAACCTTGACCATACAGAAATTCACAGCTCATGACAGAGCTGAAATTTT----- 2095
QY 795 nGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisPhePheHi 815
DB 2096 -AAACAAGTTGACTACCTCTTATTCATGGAACAGCAGATGATACCTTCATCTTCAGCA 2154
QY 815 sThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTy 835
DB 2155 GTCACTCAGATCTCCAAAGCCCTGTCGATGTTGGAGTGGATTTCCAGGCAATGTGTA 2214
QY 835 rProAsnGluArgHisSerIleArgCysProGluSerGlyGluHisTyrGluValThrLe 855
DB 2215 TACTGATGAAGACCATGGAATAGTAGCAGCAGACACCAACATATATATACCCACAT 2274
QY 855 uLeuHisPheLeuGlnGlu 861
DB 2275 GAGCCACTTCATAAAACAA 2293
RESULT 3
US-09-462-284-1
; Sequence 1, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Micheal
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5496
; TYPE: DNA
; ORGANISM: Fungus
US-09-462-284-1
Alignment Scores: 7.43e-42 Length: 5496
Pred. No.:
```

Score: 474.00 Matches: 180
Percent Similarity: 41.34% Conservative: 97
Best Local Similarity: 26.87% Mismatches: 277
Query Match: 10.20% Indels: 117
DB: 4 Gaps: 28

US-09-976-674-3 (1-863) x US-09-462-284-1 (1-5496)

QY 193 IleCysProAlaAspPro-----AlaPhe-PheSerPheIleAsnAsnSerAspLeuTr 210
Db 2365 ATATGCCAATGGAGCCCGTGGCAATACCATCGCTTTTGTTCGCCAGAGATGACCTTTA 2424
QY 210 pValAlaAsnIleGluThrGlyGluGluArgArgLeuThrPheCysHisGlnGlyLeuSe 230
Db 2425 TGTC-----TGGGATAATGGTACCGTTACTCGCATTTACT----- 2458
QY 230 rAsnValLeuAspAsp-----ProLysSerAlaGlyValAlaIleAlaThrPheValIleGL 247
Db 2459 -----GATGATGGTGGCCCGACATGTTCCACGGCGTGGCGACTGGATCTATGA 2508
QY 247 nGluGluPhe-----AspArgPheThrGlyTyTrpCysProThrAlaSerTrp-- 264
Db 2509 AGAGGATCCTCGCGCATCGCTAC---GGCTTGTGTTCTTCGCCGATGGTGAATATCT 2565
QY 265 -----GluGlySerGluGlyLeuLysThrLeuArgIleLeuTyTrpGluGluVa 280
Db 2566 GGCTTACTTGAGCTTCAATGAGACTGGGTTCGCACCTACACCGTTCCAGTAT---TATAT 2622
QY 280 lAspGluSerGluValGluValIleHisValProSerProAlaLeuGlu---GluArgLy 299
Db 2623 GGATAACCAAGACATC-----GCTCCGGCGTATCCATGGGAGCTGAA 2664
QY 299 sThrAspSerTyArgTyProArgThrGlySerLysAsnProLysIleAlaLeuLysLe 319
Db 2665 G-----ATAGGTATCCAGGTGTCCGACAGAAATCCGACCGTGGAGTTGAGTCT 2715
QY 319 uAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuVa 339
Db 2716 G-----CTTAACATCGCTAGCAAGGAGGTGAA 2742
QY 339 lGlnProPheSerSerLeuPheProLysValGluTyIleAlaArgAlaGlyTrpThrAr 359
Db 2743 GCAGGCCCGCATCGCTGCTAGAGGCTTCAACCGTGTCCAGGACCAAGAAAGGTGCT 2785
QY 359 gAspGlyLysTyArgTyAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuV 379
Db 2786 ---GGCAGGTGCTTGGCTCA-----CTGA 2808
QY 379 alLeuLeuProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlas 399
Db 2809 TACTCACACCACCGTCTGCTAGAGGCTTCAACCGTGTCCAGGACCAAGAAAGGTGCT 2868
QY 399 erAlaArgAlaValProArgAsnValGlnProTyTrpValValTyTrpGluGluValThrAsnV 419
Db 2869 CGCGGTGCATCTGCTCGCAACAG-GCTACTGTCTATCAGCGACCGATGGGACCGATG 2927
QY 419 alTrpIle---AsnValHisAspIlePheTyProPheProGlnSerGluGlyGluAspG 438
Db 2928 GATGGCTCGATAACCTTCTTCAATGAAGTATATTTGGCCCTATCAACCGTCCGACAGG 2987
QY 438 luLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyTrpLysValT 458
Db 2988 ATGCCCTACTACATGACATCTCTGAC---CATTCGGGATGGGCGCATCTGTATCTCTTC- 3043
QY 458 hrAlaValLeuLysSerGlnGlyTyArgTrpSerGluProPheSerProGlyGluAspG 478
Db 3044 -----CCCGTTTCGGGCGG----- 3058
QY 478 luPheLysCysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpClnuValLeuA 498
Db 3059 -----GAACCTATCCCATCAACAAAGGCGATGGGAGGTACG- 3097
QY 498 laArgHisGlySerLysIleTrpValAsnGluGluThrLysLeuValTyPheGlnGlyT 518

Db 3098 -----TCTATTCTGAGTATTGATCAGGAACGCCAGTTGGTGTACTACCTGTGCGA 3146
QY 518 hrLysAspThrProLeuGluHisHisLeuTyTrpValValSerTyTrpGlu-----AlaAlaG 536
Db 3147 CTCAACACACACAGCAGCGGCGCATCTCTACTCCGTCCTCTATTCACGCTTTCGGGTCA 3206
QY 536 lyGluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMetSerGlnAsn 556
Db 3207 CCGCGCTCGTCGACGACACCGTTCGCCGCTAC---TGGTCTGCTTCTTCTCCCGGACT 3263
QY 556 heAspMetPheValSerHisTyTrpSerValSerThrProCysValHisValTyTrL 576
Db 3264 CGGCTACTACATCTCATACATACGAGGCGCCAGACGTA---CCCTACAGGAACCTCTACA 3320
QY 576 ysLeuSerGlyProAspAspProLeuHisLysGlnProArgPheTrpAlaSerMetM 596
Db 3321 CGACCAACAGTAC-----AAACCACTCCCGCAATACCCGACAACG 3362
QY 596 etGluAlaAlaSerCysProProAspTyTrpValProGluIlePhePheHis----- 614
Db 3363 CCAAAGTACTCGACCAATCAAGGACTATGCATTGCCCAACATCACCTACTTCGAGCTTC 3422
QY 615 --ThrArgSerAspValArgLeuTyTrpGlyMetIleTyTrpLysProHisAlaLeuGlnProG 634
Db 3423 CCTCCCTCCGGAGAAACCTCAATGTGTGACAGCGCTTACCCCGCGGTCTCCCGCG 3482
QY 634 lyLysLysHisProThrValLeuPheValTyTrpGlyGlyProGlnValGlnLeuValAsnA 654
Db 3483 ATAGAAGTACCCCATCTTTTACCCCATACGCGCGCCAGCGCCCAAGAGTACCA 3542
QY 654 snSerPheLysGlyIleLysTy---LeuArgLeuAsnThrLeuAlaSerLeuGlyTyTrA 673
Db 3543 AGATGGCAAGCCCTGAATTTCAAGGCCTATGTCGCTCCGACAGCAATCCAGTACG 3602
QY 673 laValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaL 693
Db 3603 TAACCTGAGTGTGCACAACCGCGCACAGGTTTCAAGGACGCAAGTTCCGCTCCCGCG 3662
QY 693 euLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValA 713
Db 3663 TCAGCGCCCAACTCGGCTCTCTCGAAGACAGACAGATCTACGCGCGGAA---CAGG 3719
QY 713 laGluLysTyTrpGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyTrGlyG 733
Db 3720 CGGCCACATCCCTGGATCGATCGACACACATCGGCATCTGGGGCTGGAGTTTCGGAG 3779
QY 733 lyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
Db 3780 GCTACTTGACCAAGGTCTCTGGAGAGGACAGCGGTGCTTTCCACATTAGGAGTCACTCA 3839
QY 753 lyAlaProValThrValTrpMetAlaTyTrpGlyTyTrpThrGluArgTyTrpMetAspV 773
Db 3840 CCGCCCTGTGTGACTGGCGTTTCTACGACTCAATGTACACGAGCGCTACATGAAGA 3899
QY 773 alProGluAsnAsnGlnHisGlyTyTrpGluAlaGlySerValAlaLeuHisValGluLysL 793
Db 3900 CCTCTCGACCAATGAGGAGGCTACGAGACCGCGCGTCCGC---AAGACTGACGGGT 3956
QY 793 euProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnValHisP 813
Db 3957 TCAAGAAGCTCGAGGCGGATTTCTTGATCCAGCAGCAAGCGGCGGACGATCAACGTCCATT 4016
QY 813 hePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyTrpGlnLeuG 833
Db 4017 TCCAGAACTCGGCTGCGGTGGATCTCTGTGATGGCGATGGC----- 4060
QY 833 lnIleTyProAsnGluArgHisSer 841
Db 4061 --GTCTCTCTGAGAACCTCCATTTCG 4084


```
QY 635 sLysHisProThrValLeuPheValTyrGlyGlyProGlnValGlnLeuValAsnSe 655
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1804 GAAGTATCCCTTGGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTCTGT 1863
QY 655 rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl 673
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1864 ATTTCGTGTTAAATGGATATCTTATCTT-----GCAAGTAAGGAAGGATGGT 1911
QY 673 aValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1912 CATTCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1971
QY 693 uLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGlnGlyLeuGlnPheValAl 713
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1972 GTATCGAAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2031
QY 713 aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisClyTyrPheTyrGlyG 733
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2032 AGAA---ATGGGTTTCATTGATGAAAGAAAGAAATAGCCATATGGGCTGGCTATGGAGG 2088
QY 733 yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2089 ATAGTTTCATCACTGGCCCTTCGATCGCACTGGCTCTTTTCAATGTGGTATAGCAGT 2148
QY 753 yAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVa 773
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2149 GGCTCCAGTCTCCAGCTGGGAATATTACGCGTCTGTCTACACAGAGATTCATGGGTCT 2208
QY 773 lPro-----GluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValG 791
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2209 CCCAACAAAGGATGATAATCTTCAGCACTATAAGAAATTCAACTGTGATGCGCAAGAGCAG 2268
QY 791 uLysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVa 811
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2269 ATATTCAGAAATCTAGAC-----TATCTTCATCCAGCGGACACAGATGATAATGT 2322
QY 811 lHisPheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGl 831
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2323 GCATTTCAAACCTCAGCAGAGATGTGTAAGCTCTGGTAAATGCACAACTGGATTCCA 2382
QY 831 nLeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGluSerGly----- 848
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2383 GGCATGTGTACTCTGACCAGAACCCAGCGCTTA-----TCCGGCCTGTCCAC 2430
QY 849 -GluHisTyrGluValThrLeuLeuHisPheLeuGlnGlu 861
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2431 GAACCACTTATACCCACCATGACCCTCTCCTAAAGCAG 2470

RESULT 5
US-08-619-280A-1
; Sequence 1, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2815 Base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-619-280A-1

Alignment Scores:
Pred. No.: 5,43e-37 Length: 2815
Score: 426.00 Matches: 173
Percent Similarity: 39.08% Conservative: 106
Best Local Similarity: 24.23% Mismatches: 254
Query Match: 9.17% Indels: 182
DB: 1 Gaps: 31

US-09-976-674-3 (1-863) x US-08-619-280A-1 (1-2815)
QY 195 ProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIle 214
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 734 CCAGGAGATCCACCTTT-----CAATAACATTT-----AATGGAAGAGAAATAAATATTT 751
QY 215 GluThrGlyGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsp 234
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 752 -----CAATAACATTT-----AATGGAAGAGAAATAAATATTT 787
QY 235 AspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPheThr 254
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 788 AAT-----GGAATCCAGCTGGGTTTATGAAGAGGAATG---CTTCTCTACA 832
QY 255 GlyTyr-----TrpTrpCysProThrAlaSerTrpGluGlySerGluGlyLeuLysThr 272
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 833 AAATATGCTCTCTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 868
QY 273 LeuArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSer 292
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 869 -----TTGGCATATCGGAATTTAATGATAGATATACCATGTTATTCCTATTCCTAT 922
QY 293 ProAlaLeuGluGlu---ArgLysThrAspSerTyrArgTyrProArgThrGlySerLys 311
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 923 TATGGCGATGACAAATATCTCTAGACAAATAAATATTCATACCAAGAGCTGGAGCTAAG 982
QY 312 AsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleVal 331
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 983 AATCCGCTTCTTCGGATA-----TTTATTATTC 1009
QY 332 SerThrGlnGlnGlyLeuValGlnProPheSerSerLeuPheProLysVal----- 349
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1010 GATACCACTTACCTCGGTATGATGGTCCCGAGAGTGGCTCTTCCAGCAATGATAGCC 1069
QY 350 -----GluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAla 366
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1070 TCAAGTGATTATTATTTCAGTGGCTCAGCTGGTGTACTGATCAACGAGTATGT----- 1123
QY 367 MetPheLeuAspArgProGlnGlnTrpLeuGlnValLeuLeuProAlaLeuPhe 386
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1124 -----TTGCAGTGGCTAAAAAGAGTCCAGAAATGTTTCGGTCTGTCT 1165
QY 387 Ile-----ProSerThrGluAsn----- 392
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1166 ATATGTGACTTCAGGAAGACTGGCAGACATGGGATTTGTCCAAAGACCCAGGACATATA 1225
```

Qy	393	GlucInuArgLeu-----AlaSerAlaargAlaValProArgAsnValGlnProTyr	410
Db	1226	GAAGAAGCAGCAAGTGGATGGCGCTGGTGGATCTCTTTGTTTCAAGACAGATTTTCAGCAT	1285
Qy	411	VaiValTyrGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPhe	430
Db	1285	-----	1285
Qy	431	ProGlnSerGluGlyAlaAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGly	450
Db	1286	-----GATGCCATTTCGTACTACAAAATATTAGTCACAGGATGGC	1327
Qy	451	PheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGlu	470
Db	1328	TACAAACATATTCACTATATC-----	1348
Qy	471	ProPheSerProGlyGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThr	490
Db	1349	-----AAAGACACTGFGGAAATGCTATTCAAAATTACA	1381
Qy	491	SerGlyGluTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThr	510
Db	1382	AGTGGCAAGTGGGAGGCCATA-----AATAATATTCAGAGTAACACAGGAT	1426
Qy	511	LysLeuValTyrPheGlnGlyThr---LysAspThrProLeuGluHisHisLeuTyr---	528
Db	1427	TCACTCTTTATTTCTAGCAATGAATTTGAAGATACCCTGGGAAGAAACATCTACAGA	1486
Qy	529	-----ValValSerTyrGluAlaGlyLutIleValArgLeuThrThrProGlyPhe	546
Db	1487	ATTAGCATTTGGAGCTATCTCCAGCAAGAGTGTGT-----	1525
Qy	547	SerHisSerCysSerMet-----SerGlnAsnPheAspMet	558
Db	1526	-----ACTTGCCATCTAAGGAAGAGTCCCAATATTACACAGCAAGTTTCAGCGCAC	1579
Qy	559	PheValSerHisTyrSerSerValSerThrProProCysValHisValTyrLysLeu---	577
Db	1580	TACGCCCAAGTACTATGCATCTGCTACGCCACGCCATCCCATTTCCACCCCTTCAT	1639
Qy	578	SerGlyProAspAspAsp-ProLeuHis-----LysGlnProArgPheTrpAlaSerMe	595
Db	1640	GATGGACGCACCTGATCAAGAAATTAATCTCGGAAGAAACACAGGAATTGGAAATGCT	1699
Qy	595	tMetGluAlaAlaSerCysProProAspTyrValProProGlnIlePheHisPheHisTh	615
Db	1700	TTGAAAAAATATCCAGCTGCCT-AAAGAG-----GAATTAAGAAACTTGAAGT	1746
Qy	615	rArgSerAspValArgLeuTyrGlyWetIleTyrLysProHisAlaLeuGlnProGlyLy	635
Db	1747	AGATGAATATCTATTATGTGATCAAGATGATCTTCTCTCTCTCAA---TTTGACAGATCAAA	1803
Qy	635	sLysHisProThrValLeuPheValTyrGlyClyProGlnValGlnLeuValAsnAsnSe	655
Db	1804	GAAGTATCCTTGCTAAATCAAGTGTATGGTGGTCCCTGCAGTCAGAGTGTAAAGTCTGT	1863
Qy	655	rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl	673
Db	1864	ATTTGCTGTTAATTGGATATCTATCTT-----CCAAGTAAGGAGGATGCT	1911
Qy	673	aValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe	693
Db	1912	CATTGCTCTGGTGGATGGTGCAGAACACAGCTTTCCAAGGTGACAAACTCTCTATGCACT	1971
Qy	693	uLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAl	713
Db	1972	GTATCGAAAGCTGGGTGTTTATCAAGTGTGAGACACAGATTACAGCTGTGCAAAAATTCAT	2031
Qy	713	aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrTrpSerTyrGlyGl	733
Db	2032	AGAA---ATGGGTTTCATTGATGCAAAAAGAAATACCCATATGGGGCTGGTCTATGGAGG	2088
Qy	733	yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaGl	753

[illegible]

RESULT 6

US-08-940-391-1
 : Sequence 1, Application US/08940391
 : Patent No. 5965373
 : GENERAL INFORMATION:
 : APPLICANT: Zimmermann, Rainer; Park, John E.;
 : APPLICANT: Retig, Wolfgang; Old, Lloyd J.
 : TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION
 : TITLE OF INVENTION: PROTEIN ALPHA, AND USES THEREOF
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Felfe & Lynch
 : STREET: 805 Third Avenue
 : CITY: New York City
 : STATE: New York
 : COUNTRY: USA
 : ZIP: 10022
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
 : COMPUTER: IBM PS/2
 : OPERATING SYSTEM: PC-DOS
 : SOFTWARE: Wordperfect
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/940.391
 : FILING DATE: 01-OCT-1997
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/619,280
 : FILING DATE: 18-MARCH-1996
 : APPLICATION NUMBER: 08/230,491
 : FILING DATE: 20-APRIL-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hanson, No. 5965373man D..
 : REGISTRATION NUMBER: 30,946
 : REFERENCE/DOCKET NUMBER: LUD 5330.1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 688-9200
 : TELEFAX: (212) 838-3884
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2815 Base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : US-08-940-391-1

Alignment Scores:

Pred. No.: 5,43e-37 Length: 2815
Score: 426.00 Matches: 173
Best Similarity: 39.08% Conservative: 106
Best Local Similarity: 24.23% Mismatches: 254
Query Match: 9.17% Indels: 182
DB: 2 Gaps: 31

US-09-976-674-3 (1-863) x US-08-940-391-1 (1-2815)

```
QY 195 ProLaspProAlaPheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIle 214
DB 734 CAGGAGATCCACCTTTT-----
QY 215 GluThrGluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAsp 234
DB 752 -----CAATACATTT-----AATGAAGAGAAATAAATAATTT 787
QY 235 AspProLysSerAlaGlyValAlaThrPheValIleGlnGluPheAspArgPheThr 254
DB 788 AAT-----GCAATCCAGACTGGGTTTATGAAGAGGAATG---CTTCCTACA 832
QY 255 GlyTyr-----TrpTrpCysProThrAlaSerTrpGluGlySerGluGlyThr 272
DB 833 AATATATCTCTCTGGTGGTCTCTAATGGAAATTT-----868
QY 273 LeuArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSer 292
DB 869 -----TTGGCATATCGGAATTTAATGATAAGGATATACAGTTATGGCTATTCCTAT 922
QY 293 ProAlaLeuGluGlu-----ArgLysThrAspSerTyrArgTyrProArgThrGlySerLys 311
DB 923 TATGGCGATGAACAATATCTAGAACAAATAATATCCATACCAAGGCTGGAGCTAAG 982
QY 312 AsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleVal 331
DB 983 AATCCCTGTTTCGGATA-----TTTATTATC 1009
QY 332 SerThrGlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysVal-----349
DB 1010 GATACCACTTACCTGGTATGAGTCCCGAGGAGTGGCTGTTCCAGCAATGATAGCC 1069
QY 350 -----GluTyrIleAlaArgAlaGlyTrpThrArgAspGlyLysTyrAlaTrpAla 366
DB 1070 TCAAGTGATTATTATTCAGTTGGCTCAGCTGGGTTACTGATGAACGAGTATGT-----1123
QY 367 MetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAlaLeuPhe 386
DB 1124 -----TTGCAGTGGCTAAAGAGCTCCAGAAATGTTTCGGTCTCTCT 1165
QY 387 Ile-----ProSerThrGluAsn-----392
DB 1166 ATATGTGACTTCAGGGAAGACTGGCAGACATGGATTGTTCCAAAGACCAGGAGCATATA 1225
QY 393 GluGluGlnArgLeu-----AlaSerAlaArgAlaValProArgAsnValGlnProTyr 410
DB 1226 GAAGAAAGCAGAACTGGATGGCTGGTGGATTCTTTCTTTCAAGACCAGTTTTCAGCTAT 1285
QY 411 ValValTyrGluGluValThrAsnValTrpIleAsnValHisAspIlePheTyrProPhe 430
DB 1285 -----1285
QY 431 ProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLysThrGly 450
DB 1286 -----GATGCCATTTCGTAATAATAATATTAGTGAGCAGGATGGC 1327
QY 451 PheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGlu 470
DB 1328 TACAAACATATTCTACTATATC-----1348
QY 471 ProPheSerProGlyGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThr 490
DB 1349 -----AAAGACACTGTGGAAATATGCTATTTCAAATTACA 1381
```

```
QY 491 SerGlyGluTrpGluValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThr 510
DB 1382 AGTGGCAAGTGGGAGGCCATA-----AATATATTCAGAGTAACACAGCAT 1426
QY 511 LysLeuValTyrPheGlnGlyThr---LysAspThrProLeuGluHisHisLeuTyr---528
DB 1427 TCACTGTTTTATTCTAGCAATGAATTTGAAGAATACCTCGAAGAAGAAACATCTACAGA 1486
QY 529 -----ValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrThrProGlyPhe 546
DB 1487 ATTAGCATTTGAAGCTATCTCTCAAGCAAGAAGTGTGTT-----1525
QY 547 SerHisSerCysSerMet-----SerGlnAsnPheAspMet 558
DB 1526 -----ACTTGCCATCTAAGCAAGAAAGGTGCCAATATTACACAGCAAGTTTCAGGCAC 1579
QY 559 PheValSerHisTyrSerValSerThrProCysValHisValTyrLysLeu---577
DB 1580 TAGCCCAAGTACTATGCACCTGTCTGTACGGCCAGGCATCCCAATTTCCACCCCTTCAT 1639
QY 578 SerGlyProAspAspAsp-ProLeuHis-----LysGlnProArgPheTrpAlaSerMe 595
DB 1640 GATGGAGCGCACTGATCAAGAAATTAATAATCCTGGAAGAAACCAAGCAATTTGGAATGCT 1699
QY 595 tMetGluAlaAlaSerCysProProAspTyrValProProGluIlePheHisPheHisTh 615
DB 1700 TTGAAAAATATCCAGCTGCCT-AAAGAG-----GAAATTAAGAAACTTGAAGT 1746
QY 615 rArgSerAspValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLys 635
DB 1747 AGATGAAATTAATTTATGTTACAGATGATCTCTCTCTCTCAA---TTTGACAGATCAAA 1803
QY 635 sLysHisProThrValLeuPheValTyrGlyGlyProGlnValClnLeuValAsnAsnSe 655
DB 1804 GAAGTATCCCTTCTAATCAAGTATGTTGGTGGTCCCTGCAGTCAGAGTGTAGGTCTGT 1863
QY 655 rPhe-----LysGlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAl 673
DB 1864 ATTTGCTGTTAATTTGGATATCTTATCTTT-----GCAAGTAAGGAGGATGGT 1911
QY 673 aValValValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLe 693
DB 1912 CATTTGCTTGGTGGTGGTGGAGACAGCTTTTCCAAGGTGACAACTCTCTCTATGCACT 1971
QY 693 uLysAsnGlnMetGlyClnValGluIleGluAspGlnValClnGlyLeuGlnPheValAl 713
DB 1972 GTATCGAAAGCTGGGTGTTTATGAAGTTGAAGACCAGATTACAGCTGTCTCAAAATTCAT 2031
QY 713 aGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTrpSerTyrGlyG 733
DB 2032 AGAA---ATGGGTTTCATTGATGAAAAAGAAATAGCCATATGGGCTGGTCTCTATGGAG 2088
QY 733 yPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIleAlaG 753
DB 2089 ATACGTTTTCATCATCTGCGCCCTTCATCTCGAACTGGTCTTTTCAAAATGTGTATAGCAGT 2148
QY 753 yAlaProValThrValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspVa 773
DB 2149 GGTCTCAGCTCTCAGCTGGGAATATTACGCGTCTGTCTACACAGAGAGATTTCATGGGTCT 2208
QY 773 lPro-----GluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValG 791
DB 2209 CCCAACAAAGGATGATAATCTTCGACCATATAAGAAATTCACCTGTGTGTCACAGACAGA 2268
QY 791 uLysLeuProAsnGluProAsnArgLeuLeuLeuHisGlyPheLeuAspGluAsnVa 811
DB 2269 ATATTTTCAGAAATGTAGAC-----TATCTTCTCATCCACGGAACAGCAGATGATAATGT 2322
QY 811 lHisPhePheHisThrAsnPheLeuValSerGlnLeuLeuArgAlaClyLysProTyrG 831
DB 2323 GCACCTTTCAAAACTCAGCAGACAGATTGCTAAAGCTCTGGTTAATGTCACAGTGGATTTC 2382
```


Qy 831 nLeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGluSerGly----- 848
 Db 2383 GGCAATGGGTACTCTGACCAAGACCGGCTTA-----TCCGGCCTGTCCAC 2430
 Qy 849 -GluHisTyrGluValThrLeuLeuHisPheLeuGlnGlu 861
 Db 2431 GAACCACTATACACCCACATGACCCACTTCTCTAAAGCAG 2470
 RESULT 7
 US-08-699-103B-1/c
 ; Sequence 1, Application US/08699103B
 ; Patent No. 6107462
 ; GENERAL INFORMATION:
 ; APPLICANT: Rine, Jasper D.
 ; APPLICANT: Hampton, Randolph
 ; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
 ; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 2200 Sand Hill Road, Suite 100
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/699,103B
 ; FILING DATE: 16-AUG-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/002,581
 ; FILING DATE: 17-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Green, Grant D.
 ; REGISTRATION NUMBER: 31,259
 ; REFERENCE/DOCKET NUMBER: 09272/005001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/322-5070
 ; TELEFAX: 650/854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4982 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 US-08-699-103B-1
 Alignment Scores:
 Pred. No.: 1.5e-25 Length: 4982
 Score: 327.50 Matches: 78
 Percent Similarity: 53.99% Conservative: 37
 Best Local Similarity: 36.62% Mismatches: 83
 Query Match: 7.05% Indels: 15
 DB: 3 Gaps: 7
 US-09-976-674-3 (1-863) x US-08-699-103B-1 (1-4982)
 Qy 637 HisProThrValLeuPheValTyrGlyProGlnValGlnLeuValAsnSerPhe 656
 Db 4941 TATCCTGTATTTCCTTGGCATATGGGGACCGAATTCACCAAGTTGTCAAAACGGTT 4882
 Qy 657 Lys---GlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValVal 675
 Db 4881 TCCGTAGGATTTAATGAAGTG-----GTACCTTCACAAATTAACCGCAATGTAGTT 4831
 Qy 676 ValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsn 695
 Db 4830 GTTGTGTACGGTCGTGTGACTGCTTCAAGGTCAAGACTTTAGATCCCTTGTTCGGCAT 4771

Qy 696 GlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLys 715
 Db 4770 AGGCTCGGTATTACGAGCGCCGCGCAACCAATATCT-----GCGCTTCCTTA 4723
 Qy 716 TyrGly-----PheIleAspLeuSerArgValAlaIleHisGlyTyrPsrTyrGly 732
 Db 4722 TATGGTCTTTAACTTTTGTGTGATCCGCAAAAGATTCTCTATTATTTGGTTCATACGGG 4663
 Qy 733 GlyPheLeuSerLeuMetGlyLeuIleHisLys---ProGlnValPheLysValAlaIle 751
 Db 4662 GGGTACCTGCACACTAAAAACTTTGGAGAAAGATGCGGAAGACATTTCAAATACGGGATG 4603
 Qy 752 AlaGlyAlaProValThrValTyrMetAlaTyrAspThrGlyTyrThrGluArgTyrMet 771
 Db 4602 TCAGTTGCCAGTAACCGACTGGAGATTTACGATTTCTGTTATATCTAGAGGTACATG 4543
 Qy 772 AspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
 Db 4542 CATACTCCTCAAGAAACTTTTGATGATACGTAGTAATCAAGCGTT-----CATAATGTC 4489
 Qy 792 LysLeuProAsnGluProAsnArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnVal 811
 Db 4488 ACTGCTTTGGCACCAAGCAATAGATTTTGTGTGATGCACGGAACAGGAGATGATAACGTT 4429
 Qy 812 HisPhePheHisThr---AsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyr 830
 Db 4428 CACTTCAAAATTCCTAAAGTTCTGGACCTTTGGATCTAAATGGTGTGGAATAATAT 4369
 Qy 831 GlnLeuGlnIleTyrProAsnGluArgHisSerIleArg 843
 Db 4368 GACGTCCACGCTTTCTCCTGACTCAGATCATAGTATAAGA 4330
 RESULT 8
 US-09-229-059-1/c
 ; Sequence 1, Application US/09229059
 ; Patent No. 633172
 ; GENERAL INFORMATION:
 ; APPLICANT: Rine, Jasper D.
 ; APPLICANT: Hampton, Randolph
 ; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
 ; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 2200 Sand Hill Road, Suite 100
 ; CITY: Menlo Park
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94025
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/229,059
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/699,103
 ; FILING DATE: 16-AUG-1996
 ; APPLICATION NUMBER: 60/002,581
 ; FILING DATE: 17-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Green, Grant D.
 ; REGISTRATION NUMBER: 31,259
 ; REFERENCE/DOCKET NUMBER: 09272/005001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/322-5070
 ; TELEFAX: 650/854-0875
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4982 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-09-229-059-1

Alignment Scores:
Pred. No.: 1.5e-25 Length: 4982
Score: 327.50 Matches: 78
Percent Similarity: 53.99% Conservative: 37
Best Local Similarity: 36.62% Mismatches: 83
Query Match: 7.05% Indels: 15
DB: 4 Gaps: 7

US-09-976-674-3 (1-863) x US-09-229-059-1 (1-4982)
QY 637 HisProThrValLeuPheValTyrGlyProGlnValGlnLeuValAsnSerPhe 656
Db 4941 TATCCTCTATTTTCTTTTGCATATGGGGACCGGAATTCCTCAACAAGTTGTCAAAACGTTT 4882
QY 657 Lys---GlyIleLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValVal 675
Db 4881 TCCGTAGATTATTAAGTG-----GTAGCTTCACAAATTAACGCAATTTGTAGTT 4831
QY 676 ValIleAspGlyArgGlySerCysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsn 695
Db 4830 GTTGTTCACGCTGTGCTACTGCTTCAAAAGTCAAGACTTTAGATCCCTTGTTCGCGAT 4771
QY 696 GlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheValAlaGluLys 715
Db 4770 AGGCTCGGTGATTACGAGGCCCGGACCAATATCT-----GGCGCTTCCTTA 4723
QY 716 TyrGly-----PheIleAspLeuSerArgValAlaIleHisGlyTyrPsrTyrGly 732
Db 4722 TATGTTCTTTAACTTTTGTGATCCGCAAAAGATTTCCTTATTGCTGTCATACGGG 4663
QY 733 GlyPheLeuSerLeuMetGlyLeuIleHisLys---ProGlnValPheLysValAlaIle 751
Db 4662 GGGTACCTGACACTCAAAAACCTTTGGAGAAAGATGGCGGAAGACATTTCAAAATACGGGATG 4603
QY 752 AlaGlyAlaProValThrValTyrMetAlaTyrAspThrGlyTyrThrGluAqTyrMet 771
Db 4602 TCAGTTCGCGCAGTACCGACTGGAGATTTTACGATTCTGTTTATCTAGAGGTACATG 4543
QY 772 AspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
Db 4542 CATACTCTCAAGAAACTTTGTGATACGTAGATAACACGCTT-----CATAATGTC 4489
QY 792 LysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVal 811
Db 4488 ACTGCTTTGGCACAAGCAATAGATTTTGTGATGCACGGAACAGGAGATGATAACGTT 4429
QY 812 HisPhePheHisThr---AsnPheValSerGlnLeuIleArgAlaGlyLysProTyr 830
Db 4428 CACTTTCAAAATTCCTTAAAGTTTCTGGACCTTTTGGATCTAAATGCTGGAAAAATTAT 4369
QY 831 GlnLeuGlnIleTyrProAsnGluArgHisSerIleArg 843
Db 4368 GAGCTCCACGCTCTTCTCTGACTCAGATCATAGTATAAGA 4330

RESULT 9
US-09-221-017B-253
; Sequence 253, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
```

```
Qy 747 PheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAspThrGlyTyr 766
Db 280 TTCAAAGTCGGAGTAGCGGGCGGCTTCATAGACTCGAATCGATATGAGATTATGAC 339
Qy 767 ThrGluArgTyrMetAspValProGluAsnGlnHisGlyTyrGluAlaGlySerVal 786
Db 340 GGTAGCGCTTATTCGATGCCCGCCAGGAAATCCCGAAGGATAGCATGCTGCCAACCTG 399
Qy 787 AlaLeuHisValGluLysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPhe 806
Db 400 -----CTCAAAGCAGCGGTGATCTGAAAGGACGACATTATGCTGATTCATGGACGG 450
Qy 807 LeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAla 826
Db 451 ATCGATCCGGTCGTGGTATGCCAGCATTCACCTCTTTCTTGTGATGCTTGGCTGAAGCA 510
Qy 827 GlyLysProTyrGlnLeuGlnIleTyrProAsn 837
Db 511 CGCACCTATCTGACTCTTACGTCATCCGAGC 543

RESULT 10
US-09-221-017B-646/c
; Sequence 646, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 646:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
```

```
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...657
US-09-221-017B-646
Alignment Scores:
Pred. No.: 6.07e-24 Length: 657
Score: 299.50 Matches: 71
Percent Similarity: 50.53% Conservative: 24
Best Local Similarity: 37.77% Mismatches: 84
Query Match: 6.45% Indels: 9
DB: Gaps: 4
US-09-976-674-3 (1-863) x US-09-221-017B-646 (1-657)
Qy 605 TyrValProProGluIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMet 624
Db 566 TACAACCCGAAGGAGTTTACCACACTATCAAACTCAATCGGGCTTGAACCTGAATGCCTGG 507
Qy 625 IleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyr 644
Db 506 ATCGTGAAGCCTTATGATTTCGATCCCTCTCGCCACTATCTCTCTGATGGTACAGTAT 447
Qy 645 GlyGlyProGlnValGlnLeuVal-----AsnAsnSerPheLysGlyIleLysTyrLeu 662
Db 446 AGCGGTCCCACTCCACGAGGATTGGATCGCTATTTCATTCGATTGGGAACACTAC--- 390
Qy 663 ArgLeuAsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySer 682
Db 389 -----CTTGCAATCGAAAGTTACGTCGTGGCATGTGTGGATGGCGGTGGCACC 342
Qy 683 CysGlnArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIle 702
Db 341 GGTGCTCGCGCGAAGATGGCGAAGTGTACTACATGCAACTCGGTGATTCGAAAGC 282
Qy 703 GluAspGlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSer 722
Db 281 GATGATCAGATAGCAGCGCCACTCTATAGGACAG---CTGCCCTATGTGGATCGAGCT 225
Qy 723 ArgValAlaIleHisGlyTyrSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHis 742
Db 224 CGTATCGCATATGGGGTGGAGCTATGGCGCTATACCACTAATGAGTTGTGTGG 165
Qy 743 LysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyr 762
Db 164 GGAATGTTACATTCAAAGCGGGGATGCGCTTCTCTGGCAGACTGGCGGTTTCTTAC 105
Qy 763 AspThrGlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGlu 782
Db 104 GATTCGGTTTACACCGAAGCTTCATGCTGCTGGCAGACTGGCGGTTTCTTAC 105
Qy 783 AlaGlySerValAlaLeuHisVal 790
Db 44 ATGCTCTCT---GCTCTTGATGTG 24
RESULT 11
US-09-392-184-31/c
; Sequence 31, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392.184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 612
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: prolyl oligopeptidase)
; NAME/KEY: misc_feature
; LOCATION: (1)..(612)
; OTHER INFORMATION: n = A,T,C or G
US-09-392-184-31

Alignment Scores:
Pred. No.: 2,31e-20 Length: 612
Score: 267.00 Matches: 71
Percent Similarity: 57.59% Conservativeness: 20
Best Local Similarity: 44.94% Mismatches: 65
Query Match: 5.75% Indels: 4
DB: 4 Gaps: 0

US-09-976-674-3 (1-863) x US-09-392-184-31 (1-612)
QY 701 GluIleGluAspGlnValGluGlyLeu-GlnPheValAlaGluLysTyrGlyPheIleAs 720
Db 519 GAATGTGGGATCAGTGGGNGGACTCCCAATATCTACTCTCGAATGATTTTCATGA 460
QY 720 pLeu-SerArgValAlaIleHisGlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyL 740
Db 459 CNTAAGATCGTGTGGCAATCCNCGCTG-TCCATATGGAGGATACCTCTCCCTGATGGCAN 401
QY 740 euIleHisLysProGlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpM 760
Db 400 TAATTCAGAGTCAGATATCTACAGGTTCT-ATTTCGGGGCCCACTCACTCTGTGGA 342
QY 760 etAlaTyrAspThrGlyTyrThrGluArgTyrMetAspValProGluAsnGlnHisG 780
Db 341 TCTTCTATGAACAGGAACNCGGAGGATTANATGGTCACCTNNCCAGNATGGACAGG 282
QY 780 lyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAsnArgL 800
Db 281 GCTATTACTTAGGATCTGTGGCCATTCCANCAGGAAAGTTCCCTCTGNACCCCAATNGT 222
QY 800 euLeuIleLeuHisGlyPheLeuAspGluAsnValHisPheHisThrAsnPheLeuV 820
Db 221 TNCCTTCTTANATGGTTCTCTGGATTAGGATTTCCANTTTTCANACCAGNANANTAC 162
QY 820 alSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnGluArgH 840
Db 161 TNAGGTTTTAGTCAGGCGTGGGAACCAATGANTTNCAGAACCACTCNGGAGAAGN 102
QY 840 isSerIleArgCysProGluSerGlyGluHisTyrGluValThrLeuLeu 856
Db 101 CCAACANNAGGNTCTGAATCGGGAGGACAATATNTGACCTNNACCTT 52

RESULT 12
US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 6133012
; GENERAL INFORMATION:
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Katsuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; EARLIER FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1896

; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1896)
US-09-016-080-2

Alignment Scores:
Pred. No.: 1,5e-16 Length: 1896
Score: 240.50 Matches: 161
Percent Similarity: 35.97% Conservativeness: 107
Best Local Similarity: 21.61% Mismatches: 283
Query Match: 5.18% Indels: 195
DB: 3 Gaps: 35

US-09-976-674-3 (1-863) x US-09-016-080-2 (1-1896)
QY 132 LysArgLeuGlyValPheGlyIleThrSerTyrAspPheHisSerGlu-----Ser 148
Db 67 AAGGAAAGCGGTCTTCAAGTAACGAGATAAGCCTCAAGACGATGATTACTTCTCT 126
QY 149 GlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArgAspGlyGlyLysAsn 168
Db 127 AAGCTTTTACCTCTAC-----GATGGAAGAGG--- 153
QY 169 GlyPheMetValSerProMetLysProLeuGluIleLysThrGlnCysSerGlyProArg 188
Db 154 -----GTAAACCTTCCACCTCAGGGAACAGGATTCTAATCCAAGG 195
QY 189 MetAspProLysIleCysProAlaAspProAlaPhePheSerPheIleAsn----- 205
Db 196 TTCTCTCCA-----AATGGAAGCTTATAGCATTTACCTCAAGAGGAT 240
QY 206 -----AsnSerAspLeuTrpValAlaAsnIleGluThrGlyGluGluArg 221
Db 241 AAGGAAGAAAGGAATCAGAGCTCTACGTGATTCCACAGCATGGGAGAGGCCAGACTT 300
QY 222 LeuThrPheCysHisGlnGlyLeuSerAsnVal-----LeuAspAspProLysSerAla 239
Db 301 TTAGCAAGTTCAAATACGGGATAAAGACCTGCGCTTACCAGAGATGGGAAGATATA 360
QY 240 GlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrpCys 259
Db 361 GCGTGTGTTACCTATAGACGTTGAGAAA----- 390
QY 260 ProThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGlu 279
Db 391 -----AAAGGAATGATGACGTTCCACATTATAGG----- 420
QY 280 ValAspGluSerGluValGluValIleHisValPro-SerProAlaLeuGluGluArgLy 299
Db 421 -----GAAATACCATCTCGTTTAAATGGAG----- 445
QY 299 sThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLe 319
Db 445 ----- 445
QY 319 uAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuVa 339
Db 446 -----TTGGCTGGATCTACGGAAGAAAGAAAGCGTT 474
QY 339 lGlnProPheSerSerLeuPhePheProLysValGluTyrIleAlaArgAlaGlyTrpThrAr 359
Db 475 GTCTACCTTGTGACGTTG-----AGACGGGAAGAAAG 510
QY 359 gAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuVa 379
Db 511 AGAC-----TAACCTCAAGAAAGAACCTAAATGTTGAT-CAGATAAG 548
QY 379 lLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArg---LeuAl 398
Db 549 GTTCCACACGGTAGACTATATCTACGGCCCAAGAGGATAGGAAGAAAGAAACCTCTGAT 608
```

QY 398 aserAlaargAlaValProArgAsnValGlnProTyrValValTyrGluGluValThrAs 418
 Db 609 ATCCGATCTTTACGCTCTCGAGATGAGAAA-----GTTAGGAAGCTGACCCAGG 659
 QY 418 nvalTrpIleAsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspG1 438
 Db 660 GAAGTGG---AGGATACATCGAC---TTCCTCCCTT-----GATGACGG 698
 QY 438 uLeuCysPheLeuArgAlaAsnGluCysLysThrGlyPhe-----CysHisLeuTy 455
 Db 699 AAGCTTCGTACTTAAAGCTTACACTTTAGAAAGGGATCCCAACCAACGCCACATCTA 758
 QY 455 rLysValThrAlaValLeuLysSerGlnGlyTyrAspTrpSerGluProPheSerProG1 475
 Db 759 CCAC-----TACGATCCAA 773
 QY 475 yGluAspGluPheLysCysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpG1 495
 Db 774 GACAGGAGAACTTAAAGAGCTTCAAAAGGAT-----TTAGACAGGAACGCTTACAA 824
 QY 495 uValLeuAlaArgHisGlySerLysIleTrpValAsnGluGluThrLysLeuValTyrPh 515
 Db 825 CTCCTTA-----AACTCCGATGTCGAGGAAGTCAAGAGGCCGAGCTTGTGTACAA 875
 QY 515 eGlnGly-----ThrLysAspThrProLeuGluHisLysLeuTyrValVa 530
 Db 876 GGAGGGGTGGATCTACTATGTCGCAACGATGGCCCT---AGGCAAAACCTCTTTAGGGT 932
 QY 530 lSerTyrGluAlaIleGlyGluIleValArgLeuThrProGlyPheSerHisSerCy 550
 Db 933 CAACCTTAGAT-----GGAAGATTTAAAGGGTAATA-----GGTGAGATAGAAGCGT 980
 QY 550 sSerMetSerGlnAsnPheAspMetPheValSerHisTyrSerValSerThrProPr 570
 Db 981 TGAAGCTTCGATATAGGGATACATAGCTTCCAGCGCTCAAGATGCTGTAAACCCCAAC 1040
 QY 570 oCysValHisValTyrLysLeuSerGlyProAspAspProLeuHisLysGlnProAr 590
 Db 1041 TGAGCTGTACATATACAGG-----GATGGAAGGAGGAAGGTTACCGA 1085
 QY 590 gPhe-----TrpAlaSerMetMetGluAlaAlaSerCysProProAspTyrValProPr 608
 Db 1086 CTTTAAACAAATGGATAAAGGTTACACCTTTTCAAAA-----CC 1124
 QY 608 oGluIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMetIleTyrLysPr 628
 Db 1125 TGAACACTTTAAGCTTAAAGCAAGTACGGGGTTGAAATAGATCCCTGGGTAATGAAC 1184
 QY 628 oHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyGlyProG1 648
 Db 1185 GGTGAACCTCAGGAAGAAAGAGTATCCAGCTATTCAGAGATCCACGGTGGTCTCTAA 1244
 QY 648 nValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeu---ArgLeuAsnThrLe 667
 Db 1245 AACCGCTTAC-----GGTTACGCTTTTATGACGAGGATGCCAGCTTTT 1286
 QY 667 uAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGlnArgGlyLe 687
 Db 1287 AACCTCTAAAGGCTTCGTGATGATTTCTCAAAATCCTAGAGGAGCGATGGCTACGGAGA 1346
 QY 687 uArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValG1 707
 Db 1347 GGACTTC---CGGATATAAGGGGACACTATGGGAGGAGGATTTACAGGATTTAATGGA 1403
 QY 707 uGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleH1 727
 Db 1404 GGTAGTCGATGAACCATTAAGGAGATTGACTTTCATAGTGGGAAAGGCTAGGAGTTAC 1463
 QY 727 sGlyTrpSerTyrGlyPheLeuSer-----LeuMetClyLeuIleHisLysProG1 745
 Db 1464 CGGGGGTTCCTTGGCTGCTTACGAGAACTGTGATATGCTCGGACATACCAACAGG----- 1518
 QY 745 nValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyr----- 762

Db 1519 ----TTCAAGCCGCTGTAAACCCAGAGATCAATTTCAAATTTGGATGAAGCTTCTTCGGGAC 1574
 QY 763 ----AspThrGlyTyr-----ThrGluArgTyrMetAspValProGluAsnAsnG1 778
 Db 1575 AACGGATATAGTTATTACTTTGCTCCAGATCAAAATAGAAAAGATCCCTCGAGCAACTT 1634
 QY 778 nHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuProAsnGluProAs 798
 Db 1635 GGAAGTTATTGGGAAAAGAGC-----CCATTAAAGTACGCTCCCAACGTTGAAAC 1685
 QY 798 nArgLeuLeuLeuLeuHisGlyPheLeuAspGluAsnValHisPheHisThrAsnPh 818
 Db 1686 TCCCTGCTTATATCCACTCTACCGAGACTACAGGTGTTGGCTTCCCGAGGCAATTGCA 1745
 QY 818 eLeuValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnIleTyrProAsnG1 838
 Db 1746 ACTCTTCATATCCCTAAATACCTGGGGAAGAGAGATTGAATTGGCAATATTCCCGAGGAGA 1805
 QY 838 uArgHisSerIle 842
 Db 1806 AAATCATGACCTA 1818
 RESULT 13
 US-09-221-017B-1010/c
 ; Sequence 1010, Application US/09221017B
 ; Patent No. 644799
 ; GENERAL INFORMATION:
 ; APPLICANT: Ross, Bruce C.
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1120
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER
 ; STREET: 755 PAGE MILL ROAD
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FASTSO for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/221.017B
 ; FILING DATE: 23-DEC-1998
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1182
 ; FILING DATE: 31-DEC-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP1546
 ; FILING DATE: 30-JAN-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PP2911
 ; FILING DATE: 09-APR-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/AU98/01023
 ; FILING DATE: 10-DEC-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Monroy, Gladys H
 ; REGISTRATION NUMBER: 32,430
 ; REFERENCE/DOCKET NUMBER: 27340-20021.00
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-813-5600
 ; TELEFAX: 650-494-0792
 ; TELEX: 706141
 ; INFORMATION FOR SEQ ID NO: 1010:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3085 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular

```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3085
; US-09-221-017B-1010

Alignment Scores:
Pred. No.: 1 68e-13 Length: 3085
Score: 216.50 Matches: 139
Percent Similarity: 36.25% Conservative: 97
Best Local Similarity: 21.35% Mismatches: 262
Query Match: 4.66% Indels: 155
DB: Gaps: 21

US-09-976-674-3 (1-863) x US-09-221-017B-1010 (1-3085)
QY 318 LysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGlu 337
   ||| : : : ||||| ||||| : : : ||||| ||||| : : : ||||| : : :
Db 2888 AAGGCTCCGCTTTTCAGCTAAGCCCAAGCGGCAAGTACCTCTCATACATGGAAGGAC 2829
QY 338 -----LeuValGlnProPheSer-----Ser 344
   ||| : : : ||||| ||||| : : : ||||| ||||| : : : ||||| : : :
Db 2828 GATCGCGCAAGCCGATGCTATGTCAGGAAATTTGCCACCGGACCGCAAGCGTGCC 2769
QY 345 LeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgaspGlyLysTyrAla 364
   : : : ||| ||| ||| ||||| ||||| ||| : : : ||||| ||||| : : :
Db 2768 ATCGAAGAAAGGACCAAGTATGATCAAGGCTACGGATGGATCAACGACGACGCTCTTC 2709
QY 365 TrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeuLeuProAla 384
   : : : ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2708 TTGTGTCATG----- 2700
QY 385 LeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArgAlaValPro 404
   ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2699 -----GACAAAGAGGAGGAATGAGAACTATACCTCTTTTTCGAATATCGAGGC 2649
QY 405 ArgAsnValGlnProTyrValValTyrGluGluValThrAsnValTrpIleAsnVal--- 423
   ||| : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| |||||
Db 2648 AGCAATACCGCGCATCTACCCCTTTTGACGGAGTGAAGCTTCGATCCTCAACATGCTC 2589
QY 424 -----HisAspIlePheTyr 428
   ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2588 AAAGACGAGAGGACTACATGATCATATCCATGACAAACAAACATCCGCGATCTTCGAA 2529
QY 429 ProPheProGlnSerGluGlyGluAspGluLeuCysPheLeuArgAlaAsnGluCysLys 448
   ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2528 CCTACAAATGAATGTAGTAAACAGCGAGCTGACCCAGCTCTACGAGATAAG----- 2475
QY 449 ThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGlnGlyTyrAspTrp 468
   ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2474 -----GATCGCGGCAACCCCAATTCAGGTTACGAGTTC 2442
QY 469 SerGluProPheSer---ProGlyGluAspGluPheLysCysProIleLysGluIle 487
   : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2441 GACAAGGACGGCAACTGCGTGATACAGCGCGCTCGTAAACGGGATCGAATCCGAGTTG 2382
QY 488 -----AlaLeuThrSerGlyGluTrpGluValLeu-AlaArgHisGlySerLysII 504
   ||| : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| |||||
Db 2381 TACTACAAGGATTTGGCTACGGGCGAGTTCCGCTCTGCTGAAGAAACAC---ACTGGGAC 2325
QY 504 eTpvAlaSnGluLysLeuValTyrPheGlnGlyThrLysAspThrProLeuG1 524
   : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2324 GACACCTTCGGATCATCGGTTCACTATGCTCCAAACAAACAGAGAGCCATATGTA 2265
QY 524 uHisHisLeuTyr-----ValValSerTyrGlu----- 533
   : : : ||||| : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2264 CTGACCAACCTGGACGACGACAGACTCGTATCGTACTCTACGACCTGAAGCAGAAACAG 2205
QY 534 -----AlaAlaGlyGluLeuVal 539

```

```

Db 2204 ATCATCCGCGAGATCTTCGCCAACGAAGACTACGACGTACGCGCTCTCTCGT 2145
   ||| : : : ||||| ||||| : : : ||||| ||||| : : : ||||| : : :
QY 539 IatgLeuThrThrProGlyPheSer----- 547
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2144 AAGAGAAACTACGAATAGACCTCATGCGCTACGAAGCGGAGAACGTCGTAGTACCC 2085
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 -----HisSerCys-----SerMetSerGlnAs 555
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2084 GTAAGTGCCACCTACAAAGAGCTGCACAAGCTGATGAAAAGAAATTCAAAGGCAAGAA 2025
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 555 nPheAspMetPheValSerHisTyrSerSerValSerThrProCysValHisValty 575
   | : : : ||| : : : ||||| ||||| ||||| ||||| ||||| |||||
Db 2024 TTCTCCGTGGTCGATT-----ACGATGATGATGAGACCATCTCTGCTATCGCGTA 1974
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 575 rLys-----LeuSerGlyProAspAspProLeuHisLysLysProArgPheTr 592
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1973 CAAAGCGACAAAGCTATACGCCACTTACTACAGTTCGATA---CGCCACCAAGAACTTT 1917
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 592 pAlaSerMetMetGlu-----AlaAlaSerCysProProAspTyrValProGluI1 610
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1916 ACCCTCTCTATGACCTGATGCTCAGCTCAAG-GAGGAAGATATGCCCGAGATGCCCCC 1858
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 ePheHisPheHisThrArgSerValArgLeuTyrGlyMetIleTyrLysProHisAl 630
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1857 CATCAAAATTCAGAGCGCGCAGCGACTCATCTCCATGGCTTTTATCACTCTGCCGAAGC 1798
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 630 aLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyProGlnValG1 650
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1797 AGCCCTCGAAGGAAGAAAGTACCCCTGATCGTCAATCCGATGGAGGCCCA----- 1743
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 nLeuValAsnAsnSerPheLysGlyIleLys-----TyrLeuArgLeuAs 665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1742 -----GGCATACCGGACTCATGGGCTTCAATCCCGAGACCCA 1705
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 665 nThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGlnAr 685
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1704 GCTCTTCGCGAGCGGGATATGCCACCTGCAAGTCAATTTCCGATCTCAGCGGGATA 1645
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 gGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValIleGluAspG1 705
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1644 CGCAAGGAATCTCTCGTCCGGGATTCACACAGATCGGTGCGAAAGCCATGACGATGT 1585
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 705 nValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAl 725
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1584 GGAGGACGGTGTGCGCTAT--GCTATACGCAAGGTGGGTGGATCTCAGAGGATCGC 1528
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 725 aIleHisGlyTrpSerTyrGlyPheLeuSerLeuMetGlyLeuIleHisLysProG1 745
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1527 CATATACGGTCCGAGCACGCGTGTATGCCAGCTGATGGTCTGTGTGAAAAACCCGA 1468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 745 nValPheLysValAlaIleAlaGlyAlaProValThr----- 757
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1467 TCTCTATGCTCGGTGTGGATTACGTAGGTATCGAACATTTACACCTTCTTCGACTC 1408
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 -----ValTrpMetAlaTyrAspThrGlyTyrThrGluArgTyrMetAspValPr 774
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1407 CTTCCCAGAAATATTGGAAGCGGTTTAAAGAAATGGTCAAGGAAATTTGGTACGACCTCGA 1348
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 774 oGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGluLysLeuPr 794
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1347 -CAATCCGAGGAAGACGATATCGCAAGGAAGTGTCCCGCTTCTTCCAGATGCACAAGA 1289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 794 o-AsnGluProAsnArgLeuLeuIleHisGlyPheLeuAspGluAsnValHisPheP 814
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1288 TCAATAGCCA-----CTGTTCTGCTACAGGAGGACCAACGACCCCGGCTGAATATCA 1235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 814 heHisThrAsnPheValSerGlnLeuIleArgAlaGlyLysProTyrGlnLeuGlnI 834
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1234 ACGAGTCGATCATGATCGTAACGGCACGTGCTGCCCGGAGTTCGAAGTACCCATATGG 1175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 834 leTyrProAsnGluArgHisSerIleArgCysProGluSerGlyGluHisTyrGluValT 854
   : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
Db 1174 TGAAGTACAACGAGCGCACCGATTCCATCGTGAAGAAAACTCCATGGAGCTATACCGTG 1115
Oy 854 hrLeuLeuHisPheLeuGlnGluTyrLeu 863
Db 1114 CCAATGCTCGGTTTCTTCGCCAAACACCTG 1086

RESULT 14
US-09-280-116-172
; Sequence 172, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280.116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 172
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: prolyl oligopeptidases
US-09-280-116-172

Alignment Scores:
Pred No.: 1.61e-12 Length: 502
Score: 195.50 Matches: 47
Percent Similarity: 55.38% Conservative: 25
Best Local Similarity: 36.15% Mismatches: 54
Query Match: 4.21% Indels: 4
DB: 4 Gaps: 2

US-09-976-674-3 (1-863) x US-09-280-116-172 (1-502)
Oy 732 GlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLysValAlaIle 751
Db 1 GGTGGCTATATGTCATCAATGATCTTAAATCAGATGAAGCTTTTAAATGTGGATCC 60
Oy 752 AlaGlyAlaProValThrValTyrMetAlaTyrAspThrGlyTyrThrGluArgTyrMet 771
Db 61 GTGGTTGCACCTATCACAGACTTGAATTTGTATGCTCCTCAGCTTCTCTGAAAGATACCTT 120
Oy 772 AspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaLeuHisValGlu 791
Db 121 GGGATGCCATCTAGGAAGAAGACACTTACCAGGCGCAGCTGTCTACATAATGTTCAAT 180
Oy 792 LysLeuProAsnGluProAsnArgLeuLeuIleLeuHisGlyPheLeuAspGluAsnVal 811
Db 181 GGCITGAAGAAGAA-----AATATATATAATTCATGGAAGCTGTGACACAAAAGTT 234
Oy 812 HisPhePheHisThrAsnPheLeuValSerGlnLeuIleArgAlaGlyLysProTyrGln 831
Db 235 CATTTCACACACTCAGCAGAAATTAATCAAGCACCTAATAAAGCTGGAGTGAATTATACT 294
Oy 832 LeuGlnIleTyrProAsnGluArgHisSerIleArgCysProGlu-SerGlyGluHisTyr 851
Db 295 ATGCAGGTCTACCCCAAGTGAAGGTCAATAAC--CGTATCTGAGAAGAGCAAGTATCATCT 351
Oy 851 rGluValThrLeuLeuHisPheLeuGln 860
Db 352 CTACAGCACATCTCCAATTTCTTCAAG 379

RESULT 15
US-08-227-689-1
; Sequence 1, Application US/08227689
; Patent No. 5521081
; GENERAL INFORMATION:
; APPLICANT: Tetsuya INOKA et al.
; TITLE OF INVENTION: DNA Coding For Enzyme
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/227.689
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/917,344
; FILING DATE: July 23, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Flavobacterium meningosepticum
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: plasmid pFPEP03
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION: 1 to 259
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "promoter region"
; FEATURE:
; NAME/KEY:
; LOCATION: 260 to 316
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY:
; LOCATION: 317 to 2374
; IDENTIFICATION METHOD:
; OTHER INFORMATION: /note= "mature"
; OTHER INFORMATION: prolyl oligopeptidase coding region"
; FEATURE:
; NAME/KEY:
; LOCATION: 2375 to 2377
; IDENTIFICATION METHOD:
```


OTHER INFORMATION: /note= "stop codon"
PUBLICATION INFORMATION:

AUTHORS:

TITLE:

JOURNAL:

VOLUME:

ISSUE:

PAGES:

DATE:

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-227-689-1

Alignment Scores:

Pred. No.: 1.39e-07 Length: 2636
Score: 162.50 Matches: 124
Percent Similarity: 39.83% Conservative: 107
Best Local Similarity: 21.38% Mismatches: 224
Query Match: 3.50% Indels: 125
DB: 1 Gaps: 34

US-09-976-674-3 (1-863) x US-08-227-689-1 (1-2636)

```
QY 267 SerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluValAspGluSerGluValGlu 286
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 593 AATGATGCTCTCAGCGCAATCTGTACTTTACAGAAAGATCGCGCAGGTAAAGACGAA 652

QY 287 ValIleHisValProSerProAlaLeuGluArgLysThrAspSerTyrArgTyrPro 306
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 653 GTATTTTGTAGATCCTAAT---AAGTTTTCGGAAGAAAGAACCACTTCTCTCGCAAGTGT 709

QY 307 ArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGluPheGlnThrAspSer 326
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 710 TCTTTTAAATAAGGAAGCTCTGGTCGTATAGTATATACAGAGGAGGTTCGGACTGG 769

QY 327 GlnGlyLysIleValSerThrGlnGlu-----LysGluLeuValGln----- 340
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 770 AATAAGATTTATTCTCGATCGGAACCAACCAAAAGCACTTGATGAACTCTATTGGAT 829

QY 341 ---ProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPheArg 359
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 830 GTTAAGTTTCAGTGGAAAT-----TCATGGTTGGGA 859

QY 360 AspGlyLysTyrAlaTyrAlaMetPheLeuAspArgProGlnGlnTyrPheLeuVal 379
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 860 GATGAGGATCTTTTATCCAGCTAT---GATAAGCCAAAGAAAGGAAACCGTACTTTCC 916

QY 380 LeuLeuPro-----ProAlaLeuPhe-----IleProSerThrGluAsnGluGlu 394
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 917 GGGATGACAGATAACACAAAGTTTATTTTCATAAGTTAGGAACGAAGCAGTCTCAGGAT 976

QY 395 GlnArgLeuAlaSerAlaArgAlaValProArgAsn---ValGlnProTyrValValTyr 413
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 977 GAATTTGATTTGGGGGTGATAAATTTCCAAAGAAAGATATATAGGAGCTTATGTAAACCGAT 1036

QY 414 GluGluValThrAsnValTyrIleAsnValIleAspIlePheTyrProPheProGlnSer 433
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1037 GATCAG-----AGATATCTGGTGGTTTCGGCTGCAATGCA 1072

QY 434 GluGlyGluAspGluLeuGluCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHis 453
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1073 ACCAACGGAACGAGCT---TACATTAAAGACCTGAAGAATAAAACAGATTTTATTCG 1129

QY 454 LeuTyrLysValThrAla-----valLeuLysSerGlnGly----- 465
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1130 ATT-----ATTACAGGTTTGTATAGCAATGTAAATGTTTCAGATCCGCGGTGATPACG 1183

QY 466 ---TyrAspTyrPheSerProGlyGluAspGluPheLysCysProIleLys 484
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1184 CTTTATTTGTTCCCGNATAAGATGACCCGCAATAAGCGACTGGTAAACCAACGATTCAG 1243
```

```
QY 485 GluGluIleAlaLeuThrSerGlyGluTyr---GluValLeuAlaArgHisGlySerLys 503
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1244 AATCCAAAGCGGAACA-----TGAAAGATCTGATTGCTGAAACACCGAACA 1294

QY 504 IleTyrPValAsnGluGluThrLysLeuValTyrPheGlnGlyThr-----LysAspThr 521
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1295 TTCCAATCAAT-----ACGGGAGCGGTATTTCTTCTTCTATATATCAAGATGCA 1348

QY 522 ProLeuGluHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArg--- 540
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1349 ATCGATCAG-----GTAAGCAATATGATAAACGGAAGCTTGTAAAGGCT 1396

QY 541 LeuThrThrProGlyPheSerHisSerCysSerMet-----SerGlnAsnPheAsp 557
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1397 ATAAATTTACCGGAGTGTGTAATGCAAGCGGTTTGGGGGTGAAAAACGGAAGGAT 1456

QY 558 MetPheValSerHisTyrSerSerValSerThrProCysValHisValTyrLysLeu 577
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1457 CTGTATTACTCT---TTCACCAATATATTACGCGCAACGATCTTTAAATATAATGTA 1513

QY 578 SerGlyProAspAspProLeuHisLysGlnProArgPheTyrAlaSerMetMetGlu 597
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1514 ACAACAGGTAACTCTGAA---GTTTACCAGAGCCGGAAGTGAAGTTCAAT----- 1561

QY 598 AlaAlaSerCysProAspTyrValProGluIlePheHisPheHisThrArgSer 617
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1562 -----CGGAAAAATTATGTTTCGAGCAGGTA-----TTCATATCTTCTATCT 1603

QY 618 Asp---ValArgLeuTyrGlyMetIleTyrLysProHisAlaLeuGlnProGlyLysLys 636
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1604 GACGGGACTAAGATTCGATGATCAGCTACAGAAAGCCCTGAAAAAGACGGTAAA 1663

QY 637 HisProThrValLeuPheValTyrGlyGly-----ProGlnValGln 650
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1664 AACCTACATATATATACAGCTACGGAGGATTTAATATCAGCTCTTCAGCCTGCTTCTCT 1723

QY 651 LeuValAsnAsnSerPheLys-----GlyIleLysTyrLeuArgLeuAsnThrLeu 667
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1724 GTTGTAAATCAATCTCGATGGAAACGGTGGTATT----- 1759

QY 668 AlaSerLeuGlyTyrAlaValValValIleAspGlyArgGlySerCysGlnArgGlyLeu 687
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1760 -----TATGCTGTTCGCAATATC-----CGTGGTGGTGGAGAAATATGGTAAAG 1801

QY 688 ArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGlu 707
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1802 AATGGCATGATCCCGGAACATAATCCAGAAAAAGATGATTTATTAATGACTTTATTGCA 1861

QY 708 GlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHis 727
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1862 GCCGGAGAGTACTTA---CAGAAAAACGGTTATACATCTAAGGAATATATGGCGTTTCC 1918

QY 728 GlyTrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPhe 747
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1919 GGACGTTCACACGGCGGTCTTCTGTAGGGGCTACGATGACAATGCGCCCTGATTTGGCA 1978

QY 748 LysValAlaIleAlaGlyAlaProVal----- 756
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 1979 AAGTTGCAATCCCGGAGTAGGAGTACTGGATATGCTTCTTATATAAGTTTACAGCT 2038

QY 757 ---ThrValTrpMetAlaTyrAspThrGly----- 765
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 2039 GGTGCGGTTTGG---GCTTATGATTACGGTACAGCAGACAGCAAGAGAAATGTTTGA 2095

QY 766 TyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySer 785
      :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
Db 2096 TACCTGAAGTCTTATTCTCCGGTA-----CATAACGTAAAAAGCGGAAC 2140
```

Search completed: December 12, 2002, 13:55:08
Job time : 103 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 12, 2002, 12:09:28 ; Search time 1947 Seconds

(without alignments)
7178.585 Million cell updates/sec

Title: US-09-976-674-3

Perfect score: 4646

Sequence: 1 MATTGTPTRDRCDAATDDP.....CPESGHEVTLHLPIQEVYL 863

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_p2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US09976674/runatc_04122002_162531_7145/app-query.fasta.1.1031
-DB=EST -GFM=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pio -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09976674.ecgn_1.1.1391.etruncat_04122002_162531_7145 -MCPV=6 -ICPV=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inu:*
20: em_gss_pln:*
21: em_gss_vrl:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pio:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2887.5	62.2	3143	11 AK016546	AK016546 Mus muscu
2	1638	35.3	1137	11 BM553230	BM553230 AGENCOURT
3	1602	34.5	1043	13 B0051605	B0051605 AGENCOURT
4	1516	32.6	921	14 B0877413	B0877413 AGENCOURT
5	1512.5	32.6	932	14 B0684956	B0684956 AGENCOURT
6	1486.5	32.0	982	14 B0678015	B0678015 AGENCOURT
7	1468	31.6	880	14 B0897707	B0897707 AGENCOURT
8	1441	31.0	879	14 B08895618	B08895618 AGENCOURT
9	1439	31.0	879	14 B0889360	B0889360 AGENCOURT
10	1432	30.8	866	13 BM461814	BM461814 AGENCOURT
11	1410	30.3	872	14 B0689588	B0689588 AGENCOURT
12	1403	30.2	910	14 B0949519	B0949519 AGENCOURT
13	1395	30.0	1051	14 BM915935	BM915935 AGENCOURT
14	1383	29.8	821	13 B1690774	B1690774 603314392
15	1372.5	29.5	902	14 B0642814	B0642814 AGENCOURT
16	1369	29.5	787	13 B1223668	B1223668 602943581
17	1360.5	29.3	938	13 B0685669	B0685669 AGENCOURT
18	1320.5	28.4	1030	13 BM424091	BM424091 AGENCOURT
19	1308	28.2	1195	13 BM556123	BM556123 AGENCOURT
20	1292.5	27.8	974	13 B06423964	B06423964 AGENCOURT
21	1274.5	27.4	975	14 B0681307	B0681307 AGENCOURT
22	1272	27.4	1223	14 BM809492	BM809492 AGENCOURT
23	1259.5	27.3	896	13 BG921463	BG921463 602824365
24	1246	26.8	747	13 B1258117	B1258117 602970350
25	1226.5	26.4	997	14 B0642151	B0642151 AGENCOURT
26	1207	26.0	880	14 B0681942	B0681942 AGENCOURT
27	1195	25.7	822	13 BG330712	BG330712 602430226
28	1194.5	25.7	785	13 B1148072	B1148072 602912419
29	1188	25.6	727	13 B1855677	B1855677 603383205
30	1177	25.3	698	12 BG108176	BG108176 602280402
31	1169	25.2	1032	12 BG167854	BG167854 602339821
32	1165.5	25.1	1023	14 B0681703	B0681703 AGENCOURT
33	1164	25.1	809	13 BM045258	BM045258 603623210
34	1154.5	24.8	1080	12 BF203783	BF203783 601868764
35	1153	24.8	1042	13 BM557438	BM557438 AGENCOURT
36	1133	24.4	908	12 BG767446	BG767446 602741246
37	1129.5	24.3	828	12 BG424090	BG424090 602448161
38	1123	24.2	725	13 BM018261	BM018261 603645911
39	1115	24.0	881	12 BF535546	BF535546 602050541
40	1110	23.9	668	12 BG290429	BG290429 602308277
41	1094.5	23.6	1077	13 BM560972	BM560972 AGENCOURT
42	1094	23.5	892	13 B1160407	B1160407 602864430
43	1089	23.4	798	12 BG178503	BG178503 602328312
44	1081	23.3	931	12 BG421210	BG421210 602451701
45	1078	23.2	1088	13 BM458946	BM458946 AGENCOURT

ALIGNMENTS

RESULT 1
AK016546
LOCUS
DEFINITION AK016546 3143 bp mRNA linear HTC 19-JAN-2002
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4932434F09;homolog to DIPEPTIDYL PEPTIDASE 8, full
insert sequence.
ACCESSION AK016546
VERSION AK016546.1 GI:12855334
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (Strain: C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:4932434F09.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

Fri Dec 13 12:15:17 2002

us-09-976-674-3.rst

Page 2

REFERENCE	TITLE	JOURNAL	MEDLINE	PUBMED
1	Carninci, P. and Hayashizaki, Y.			
	High-efficiency full-length cDNA cloning			
	Meth. Enzymol. 303, 19-44 (1999)			
2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
	Genome Res. 10 (10), 1617-1630 (2000)			
3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
	Riken integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichipillary sequencer			
	Genome Res. 10 (11), 1757-1771 (2000)			
4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arkawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Konno, S., Yamanka, I., Salto, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Salto, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischiemann, W., Gaasterland, T., Gissi, C., King, B., Kochava, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Resole, G., Quackenbush, J., Schriml, L.M., Stambli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotmann, M., Hume, D.A., Kamuya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, R., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seyer, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L., Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsukki, S. and Hayashizaki, Y.			
	Functional annotation of a full-length mouse cDNA collection			
	Nature 409 (6821), 685-690 (2001)			
5	(bases 1 to 3143)			
	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arkawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hizoaka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
	Direct Submission			
	Submitted (10-JUL-2000) yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp, Tel: 81-45-503-9222,			

COMMENT

Fax:81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAAGAAAGATCCAAAGACCTCTTTTCTTTTTTNN 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAAGAGCAATTCGACGTAAATTGAATAATCACCCTCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from lambda f1c I. Cloning sites, 5' end: SalI; 3' end: BamBI. Host: DH10B.

FEATURES

SOURCE

Location/Organism
1. .3143 /organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:4932434F09"
/db_xref="MGI:MGI:1907554"
/db_xref="taxon:10090"
/clone="4932434F09"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
354..3005
/note="data source:SPFR, source key:Q9HBM5, evidence:ISS homolog to dipeptidyl peptidase 8 putative"
/codon_start=1
/protein_id="BAB30295.1"
/db_xref="GI:12855335"
/db_xref="MGI:MGI:1921638"
/translation="MAAAMEEQLGVEIEFTAECEBNGESODRPLKEPFYERYSMS QLKKLADTRKYGHYMMAKAPHPFMFKRIDPDRPSDRLYYLAMSGENRENTLFYS IPTINAAVALMLSMKRLIDFOATLDGYGSSEELRRKRKIGVGLAAVDYHGSG GLTFPGASGVITIKGGPDLFPQOPLRPILVTSCPNIRMPDKLRPPADPAIFHS MDIVISLVLTREERRITYNHNELANMEEDPRSGAVTVLYOEFDKISGYWCPOAER TPGGGLIRLILEENDESEVELIHVTSPLMETRAASFRTPKGTANPVTEKMSIV VDAAGGIIDVIDELVOPEILFEGEVYLARAQTPBGKAWSILDRSQTHQIVLI SPELFIIVEDDAMDROLRLISVPDSVPTLAIRGETWDINIHDIFEPORHEDEIE FTFASECKTGFRHLRYKTLSLKSKYKRSSGGLPAPSDFCPIKEETITSGEVIAG RHSGNIWMDARLIVYPEGRKDPLEHHLYVTSVANPGEVRLTDRTGYSHCSHC DPTISKSNOKNPHCVSLKYRKSSPEDDPNHTKFGVATITLDSAGPLPDVTPEITSFE STTGFTSLYGMVLVRPHDLPGRKYPVALVFYIGGVOVLVNNRKGVAFPRFLNTLASIG YVVVINDRSGSCHRGLEFGAKFYKMGOIELDDOVEGLQYLASQDFIDLDRVGHWMS YGVALISALMALMOSSDIFFRVAICAGPATLMIFDTGYTERYMGHPDNOEGUYGSVAM QAEPKPEPBRLLILGHGFENDVHPAHTSIILSFVRACKPYDLOLYIQEHRSHRVPE SGEHYEHLLHLYIOENIGSRILALKYI"

BASE COUNT

894 a	688 c	773 g	788 t
-------	-------	-------	-------

ORIGIN

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 2887.50	3143	520	139	197	3	
Best local Similarity: 76.72%						
Best local Similarity: 60.54%						
Query Match: 62.15%						

US-09-976-674-3 (1-863) x AK016546 (1-3143)

Oy 6 ThrProThraAlaAspArgGlyAspAlaAlaLeuThrAspAspProValalaIatg---Phe 24
||| :::: |||::: ||| ||| |||
Db 402 ACTGCAGAGGTGTGAGGAGCAATATGGTAATCATACAGACCGCACCAAGTCGACCTTTT 461
25 GluValGlnIshSerTrpAspGlyLeuArgSerIleIleHisGlySerArgIysTyr 44

```

Db 462 TATGTCGAACGGTATTTCTTGAGTCAGCTGAAAGAGCTGCTGCGATACCAAGAAATATC 521
Qy 45 SerGlyLeuIleValAsnLysAlaProHisAspPheGlnPheValGlnLysThrAspGlu 64
Db 522 CACGGCACTAAGATGAGTAAAGCGCCACATGACTTATGTTGTAAGAGACGCGATCCA 581
Qy 65 SerGlyProHisSerHisArgLeuTyrThrLeuGlyMetProTyrGlySerArgGluAsn 84
Db 582 GATAGGCTCTCCTCAGACAGAGGCTCATTAACCTTGCCATGCTGTGGGAGACAGAGAAAT 641
Qy 85 SerLeuLeuTyrSerGluIleProLysLysValArgLysGluAlaLeuLeuLeuSer 104
Db 642 ACACATATTTATCTGAAATCCCTAAACCATCAACAGACAGACAGCTTATATGCTTCC 701
Qy 105 TrpLysGlnMetLeuAsnHisPheGlnAlaThrProHisHisGlyValTyrSerArgGlu 124
Db 702 TGGAAAGCCCTTTGGATCTTTTCAGCGACACTAGACTAGACTATGGATGATTTCTCGAGAG 761
Qy 125 GluGluLeuLeuArgGluArgLysArgLeuGlyValPheGlyIleThrSerTyrAspPhe 144
Db 762 GAAGAGCTACTCAGAGAAAGAAAGCCGCTTGAGACCGTGGGAGATCGCAGCTTATGATATAC 821
Qy 145 HisSerGlySerGlyLeuPheLeuPheGlnAlaSerAsnSerLeuPheHisCysArgAsp 164
Db 822 CACCAGAGAACGGCAACTTCCTGTTTCACACTGCGACGCGCATTTACCATATCAAGAT 881
Qy 165 GlyGlyLysAsnGlyPheMetValSerProMetLysProLeuGluIleLysThrGlnCys 184
Db 882 GGAGACCCACATGATTTTACACAAACAGCCCTTGCGGCCAATTTAGTGAAGACCTAGTGT 941
Qy 185 SerGlyProArgMetAspProLysIleCysProAlaAspProAlaPhePheSerPheIle 204
Db 942 CCAATATATACAAATGAGATTCAAAATATATGCTGCTGATCCAGACTGATAGCTTTCATT 1001
Qy 205 AsnAsnSerAspLeuTyrValAlaAsnIleGluThrGlyGluGluArgArgLeuThrPhe 224
Db 1002 CACAGCAATGATATTTGGATATCAAACTCGTAACCGAGAAAGACGAGATCATCAATAC 1061
Qy 225 CysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAlaGlyValAlaThrPhe 244
Db 1062 GTACACATGAGCTAGCCCAACATCGAGAAAGATCCCAATCAGCTGCGTCCACCTTT 1121
Qy 245 ValIleGlnGluGluPheAspArgPheThrGlyTyrTyrProTyrCysProThrAlaSerTrp 264
Db 1122 GTCTCTCAGAAAGATTTTGACAGATCTGCTGCTAGCTGCTGCTGCCAAACGGAAGA 1181
Qy 265 GluGlySerGluGlyLeuLysThrLeuArgIleLeuTyrGluGluValAspLysSerGlu 284
Db 1182 ACTGCTAGTGGTGT---AAATCTTGAATTTCTCTATGAAGAAATGATGATCATCTGAG 1238
Qy 285 ValGluValIleHisValProSerProAlaLeuGluGluArgLysThrAspSerTyrArg 304
Db 1239 GTGGAGATTAATTCATGTTACGCTCCCATGTTGTAACAGAGAGGCGAGATTTCTTCTG 1298
Qy 305 TyrProArgThrGlySerLysAsnProLysIleAlaLeuLysIleuAlaGluPheGlnThr 324
Db 1299 TATCCCAAAACAGCAGCAACCAACCAAGGCTCACTTCAAGATGTGCGAGATTGTTGTT 1358
Qy 325 AspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGluProPheSerSer 344
Db 1359 GATGCTGACGAGGAGATATATGATGTATAGATTAAGAAAGCGTTCAACCTTTCAGAGATT 1418
Qy 345 LeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPthrArgAspGlyLysTyrAla 364
Db 1419 CTGTTTGGAGGATGGAATATATGCTCCAGACCGGATGAGACTCCAGAGAGAAACATGCC 1478
Qy 365 TrpAlaMetPheLeuAsnArgProGlnGlnIlePheGlnLeuValLeuLeuProProAla 384
Db 1479 TGGTCCATCTCACTAGACCGTTCCAGACTACCTGCGAGATATGTTCTGATCTCCCTGAG 1538
Qy 385 LeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArgAlaValPro 404
Db 1539 TTAATCATCCAGTAGAAGATGATGCCATGACAGACAGAGACTTATAGACTACTCTCT 1598
Qy 405 ArgAsnValGlnProTyrValValTyrGluGluValThrAsnValThrPheAsnValHis 424
Db 1599 GACTGTGACACACACATGATCATATGAGAAACAAACAGACATCTGATTAATATCCAC 1658
Qy 425 AspIlePheTyrProProGlnSerGlnGlyLysGluAspGluLeuCysPheLeuArgAla 444
Db 1659 GATATTTTCACTGTTTTCCTCAAACTCAT---GAAGATGAAGATGATTTATTTTTCCTC 1715
Qy 445 AsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeuLysSerGln 464
Db 1716 TCTGAATCAAAACAGGTTTTCGTCATCTGTATAAATCAATCATTTTAAAGAGAGAC 1775
Qy 465 GlyTyrAspTrpSerGlnProPheSerProGlyGluAspGluPheLysCysProIleLys 484
Db 1776 AAATATAAAGCGTCCAGTGGTGGAGACTACTGCCCCCAAGTGAATTCATGATCCATCAAA 1835
Qy 485 GluGluIleAlaLeuThrSerGlyGluTyrPheGluValLeuAlaArgHisGlySerLysIle 504
Db 1836 GAAGAAATAAACATTAATCAACGATGCTAATGGAAGATCTTGGCCGCGATGATTAATATC 1895
Qy 505 TrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAspThrProLeuGlu 524
Db 1896 TGGGTGATGAAGCCAGAAAGCTGGTGTACTTGAAGCCACCAAGACTCTTCCTTGGAG 1955
Qy 525 HisHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArgLeuThrThrPro 544
Db 1956 CATCACTGTACGTGACCAAGTATGCAAAACCTGGAGAGAGTGTCAGGCTGACTACCGCT 2015
Qy 545 GlyPheSerHisSerCysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSer 564
Db 2016 GGCTACTACACTCTGCTGCTGCTGACCGCGCTTGTGACTTCTTATAGTAGATGAC 2075
Qy 565 SerValSerThrProProCysValHisValTyrLysLeuSerGlyProAspAspAspPro 584
Db 2076 AACCAAGAAATCCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2135
Qy 585 LeuHisLysGlnProArgPheThrAlaSerMetGluAlaAlaSerCysProProAsp 604
Db 2136 GTTCATTAACAAAGAAATTTTGGGCCACCAATTTGGATTCAGCAGGCTCTCTCTGAC 2195
Qy 605 TyrValProProGlnIlePheHisPheHisThrArgSerAspValArgLeuTyrGlyMet 624
Db 2196 TACACCTCTCCGAAATTTTCTTTTGAATACTACTGATGATTAACCTGATGGAATG 2255
Qy 625 IleTyrLysProHisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyr 644
Db 2256 TTGTTAAGCCCTCAGACTGACCTCAACCTGGAAGAAATACCCACGCTGTTATCATATAT 2315
Qy 645 GlyGlyProGlnValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArgLeu 664
Db 2316 GGTGTCTCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2375
Qy 665 AsnThrLeuAlaSerLeuGlyTyrAlaValValIleAspGlyArgGlySerCysGln 684
Db 2376 AACACCTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2435
Qy 685 ArgGlyLeuArgPheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAsp 704
Db 2436 CGAGAGCTTAATTTGAAGGCGCTTTAAATTAATAAATGAAGGCTCAAAATGAATCATGAT 2495
Qy 705 GlnValGluGlyLeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgVal 724
Db 2496 CAAGTGAAGGACTCAGTACCTACATCTCAGTATGACTTATGACTTATGACTTATGACT 2555
Qy 725 AlaIleHisGlyTyrSerTyrGlyLysPheLeuSerLeuMetGlyLeuIleHisTyrPro 744
Db 2556 GGCATCCACGGCTGCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2615
Qy 745 GlnValPheLysValAlaIleAlaGlyAlaProValThrValTrpMetAlaTyrAspThr 764
Db 2616 GATATCTTCCGGGTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2675

```

QY 765 GlyTyrThrGluArgTyrMetAspValProGluAsnAsnGlnHisGlyTyrGluAlaGly 784
|||||
Db 2676 GGATACACGAGCGCTATATGCTGACCTGACCCAGATGAAAGAGGCTACTACTAGGA 2735
QY 785 SerValAlaLeuHisValGluLeuProAsnGluProAsnAlaGluLeuHis 804
|||||
Db 2736 TCTGTGGCCATGAGGAGGAGGAGTTCCTCCAGAACCCGGTATCTCTTCAT 2795
QY 805 GlyPheLeuAspGluAsnValHisPhePheHisThrAsnPheLeuValSerGlnLeu 824
|||||
Db 2796 GGGTCTTGATGAGATGTTCACTTTCACACACACAGATATGCTGAGTTTAAAGT 2855
QY 825 ArgAlaGlyProTyrGlnLeuGlnLeuPheProAsnGluArgHisSerIleArgCys 844
|||||
Db 2856 AGGCGCTGAAAGCATATGACTTACAGATCTATCTCAGAGAGGACAGCATCAGATT 2915
QY 845 ProGluSerGlyGluHisTyrGluValThrLeuLeuHisPheLeuGlnGluTyrLeu 863
|||||
Db 2916 CCTGAGTCTGAGAACACTATGAACTGACCTGCTCCACCTTACAGAGAACCTT 2972
RESULT 2
BM553230 1137 bp mRNA linear EST 20-FEB-2002
LOCUS BM553230
DEFINITION AGENCOURT_6572641 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467115
ACCESSION 5', mRNA sequence.
VERSION BM553230
KEYWORDS BM553230.1 GI:18791777
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1137)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM1968 row: j column: 12
High quality sequence stop: 672.
Location/Qualifiers
FEATURES
source
1. 1137
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5467115"
/clone_1ib="NIH_MGC_41"
/issue_type="amelanotic melanoma, cell line"
/lab_host="DH10b (phage-resistant)"
/note="Organ: skin; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

BASE COUNT 253 a 346 c 320 g 216 t 2.others

ALIGNMENT SCORES:

Pred. No.: 2.05e-185 Length: 1137
Score: 1638.00 Matches: 338
Percent Similarity: 89.56% Conservative: 5
Best Local Similarity: 88.25% Mismatches: 22
Query Match: 35.26% Indels: 19

DB: 13 Gaps: 7
US-09-976-674-3 (1-863) x BM553230 (1-1137)
QY 114 AlaThrProHisHisGlyValTyrSerArgGluGluLeuLeuArgGluArg 133
|||||
Db 3 GCCAGGCCCCACCACTGGGGTCTTACTCTCTGAGAGAGAGCTGCTGAGAGCGGAAACGC 62
QY 134 LeuGlyValPheGlyIleThrSerTyrAspPheHisSerGluSerGlyLeuPhe 153
|||||
Db 63 CTGGGGCTCTTGGCATCACCCTCTACGACTTCACAGCCAGAGATGGCTCTTCTTTC 122
QY 154 GlnAlaSerAsnSerLeuPheHisCysArgAspGlyGlyAsnGlyPheValSer 173
|||||
Db 123 CAGGCGAGCAACACCTCTTCCACTGCGGAGCGGCAAGAACGGCTCATAGTGTCC 182
QY 174 PrometLysProLeuGluIleLysThrGlnCysSerLysProArgMetAspProLysIle 193
|||||
Db 183 CCTATGAAACCGCTGGAAATCAAGACCACTGCTCAGAGGCCCGGATGACCCCAAAATC 242
QY 194 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrPvalAlaAsn 213
|||||
Db 243 TGGCTTCCGACCTGCTGCTTCTCTTCTATCATATACAGCCAGCTGTGGTGCCAC 302
QY 214 IleGluThrGlyGluArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 233
|||||
Db 303 ATGAGACAGGCGAGGAGCGGCGCTGACCTTCCACCAAGGTATATCAATGTCCTG 362
QY 234 AspAspProLysSerAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPhe 253
|||||
Db 363 GATGACCCCAAGTCTGGGGTGTGGCCACTTCTCTCAATACAGAAAGGTTGACCCCTTC 422
QY 254 ThrGlyTyrTrpCysProThrAlaSerTrpGluIleSerGluGlyLeuValThrLeu 273
|||||
Db 423 ACTGGTACGAGTGTGTCCTCCACAGCTCTCTGGAAAGTTCAAGAGGCCCTCAAGACCTG 482
QY 274 ArgIleLeuTyrGluGluValAspGluSerGluValGluValIleHisValProSerPro 293
|||||
Db 483 CGAATCCTGATGAGAAAGTCGATGATCGAGGTGAGGTGATTCACGCTCCCTCTCT 542
QY 294 AlaLeuGluGluArgLysThrAspSerTyrArgTyrProArgTrpGlySerLysAspPro 313
|||||
Db 543 GCGCTAAGAAAGAAAGAGAGAGAGCTGATCGGACCCAGAGAGGACAGCAAGAAATCC 602
QY 314 LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 333
|||||
Db 603 AAGTTCCTTGAACATCGGTGATTCACAGTACAGCCAGGCAAGATCGTCTGACC 662
QY 334 GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluTyrIleAla 353
|||||
Db 663 CAGAGCAAGGAGCTGCTGAGCCCTTCCAGCTCTTCCGAAAGGTGGAGTACATCGCC 722
QY 354 ArgAlaGlyTrpPheArgAspGlyLysTyrAlaThrPheLeuPheLeuAspArgProGln 373
|||||
Db 723 AGGCGCGGTGAGCCCGGAGTGGCAATAGCGCTGGCCATGTTCTCGAGCCGAGGCCAG 782
QY 374 GlnTrpLeuGlnLeuValLeuLeuProAlaLeuPheIleProSerTrpGluAsnGlu 393
|||||
Db 783 CAGTGGCTTACGCTGCTGCTCTCCCGCCGCG -CTGTTTATCCCGACACAGAGAAAGAG 841
QY 394 GlnGlnArgLeuAlaSerAlaArgAlaValProArgAsnValGlnProTyrValTyr 413
|||||
Db 842 GAGCAGCGCTACCTCTGCGCAGAGGTGTCCAGAAATGCCAGCGGTATGTGTGTAC 901
QY 414 Glu-GluValThrAsnValTrp-IleAsnValHis-Asp-IlePheTyrProPheProGln 432
|||||
Db 902 GAGGAGGTCAACCACTCTGGATCATATGTTCATGAGCAATCTTATCTTCTCCGCC 961
QY 432 nserGluGlyGlu---AspGluLeuCys-PheLeuArgAlaAsnGluCys---LysThr 450
|||||
Db 962 CATCAAGGAGAAAGACCAAGCTTCTGCTTCTCCCGGCAAGNAATCAAGAAACGG 1021
QY 450 Lyphe-CysHisLeuTyrLysValThrAlaValLeu-----LysSerGlnGly 465

```

||||| ||||| ||||| ||||| |||||
Db 1022 GCTTCGCGCCATTTT-----GTAACAAAGTCCACGCGCGTTTAAATAATCCAGG 1075
OY 466 ---Tyrasptpserglu-----PropheserProgluValaspGluphelsCys 481
Db 1076 GCTTACCCAAATGGAGGAGGAGCCCTTCAACCCCGCGGAGAAA--AAATGT 1129

RESULT 3
BO051605
LOCUS
DEFINITION
ABENCOURT_6866328 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5932532
5', mRNA sequence.
BO051605
VERSION
BO051605.1 GI:19810945
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1043)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DMS/NCI
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
plate: LICM215 row: j column: 21
High quality sequence. Stop: 641.
Location/Qualifiers
1. 1043
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5932532"
/clone_lib="NIH_MGC_106"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: blood; Vector: pORF7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 215 a 334 c 296 g 191 t 7 others
ORIGIN

Alignment Scores:
Pred. No.: 3,76e-181 Length: 1043
Score: 1602.00 Matches: 312
Percent Similarity: 94.26% Conservative: 0
Best Local Similarity: 94.26% Mismatches: 14
Query Match: 34.48% Indels: 5
DB: 14 Gaps: 2

US-09-976-674-3 (1-863) x BO051605 (1-1043)

OY 58 PheValGlnLysThrAspGluSerGlyProHisSerHisArgLeuTyrTyrLeuGlyMet 77
|||||
Db 1 TTTGTGCGAAGAGGAGTGTGGGCCCATCCACCCCTTACTACTGGGAATG 60
OY 78 ProTyrGlySerArgGlnAsnSerLeuLeuTyrSerGluIleProLysValAlaArgLys 97
|||||
Db 61 CCAATGTCACCGGAGACACTCCCTCTTACTGTGATGCCAAGAGGTCGGGAAA 120
OY 98 GluAlaLeuLeuLeuLeuSerTyrLysGlnMetLeuAspHisPheGlnAlaThrProHis 117
|||||

```

```

Db 121 GAGGCTGTGTCCTCTGCTGCGAGACAGATGCTGATCATTTCCAGCCACGCCCCAC 180
OY 118 HisGlyValTyrSerArgGluGluLeuLeuArgGluValArgLysArgLeuGlyAlaPhe 137
|||||
Db 181 CATGGGGTCTACTCTCGGAGAGAGAGCTGTGAGGAGCGGAAACGCTGGGGTCTTC 240
OY 138 GlyIleThrSerTyrAspPheHisSerGluSerGlyLeuPheLeuPheGlnAlaSerAsn 157
|||||
Db 241 GGCAATCACTCTCTAGACACTTCCACAGCAGAGAGTGGCTTCTCTTCCAGACCGACAC 300
OY 158 SerLeuPheHisCysArgAspGlyLysLysAsnGlyPheMetValSerProMetLysPro 177
Db 301 AGCCTCTTCCACTGCGCGGAGCGGCAAGAGCGCTTCAATGATGTCCTCATGAAACCG 360
OY 178 LeuGluIleLysThrGlnCysSerGlyProArgMetAspProLysIleCysProAlaAsp 197
Db 361 CTGGAATCAAGACCCCACTGCTCAGAGGCGCGGATGAGACCCCAAAATGTGCTCCAC 420
OY 198 ProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrPValAlaAsnIleGluThrGly 217
Db 421 CCTGCTTCTTCTCTCTTATATATACAGCAGCTGTGGTGCCACATGACAGACAGC 480
OY 218 GluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLys 237
Db 481 GAGAGCGCGCGGCTGACCTTCTGCCACCAAGGTTATCATATGCTGTGATGACCCCAAG 540
OY 238 SerAlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrp 257
Db 541 TCTGGGGGTGTGGCCACCTTGTGTCATACAGAAAGTTTCAGCCGTCCTACTGGGTACTGG 600
OY 258 TrpCysProThrAlaSerTyrGlnGlySerGluGlyLeuLysThrLeuArgTlleuTyr 277
Db 601 TGTGTCCCAACAGCTCTCTGGGAGAGTTCAAGAGGCTTCAAGACCTCGAATCTGTAT 660
OY 278 GluGluValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluGlu 297
Db 661 GAGGAAGTCGATGAGTCGAGGTGAGGTCAATTCAGTCCCTCTCGCGGTAGAAAGA 720
OY 298 ArgLysThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeu 317
Db 721 AGGAAGAGGAGCTCTATCGGTACCCAGACAGGACACAGAAATCCCAATATGCTTG 780
OY 318 LysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGlu 337
Db 781 AAACGTGCTNAGTCCACATGACAGCAGGCAAGATCGTGTGACCCAGAGAGAGAG 840
OY 338 LeuValGlnProPheSerSerLeuPheProLysValGluTyrIleAla-ArgAla-GlyT 357
Db 841 CTGGTGCAGCCCTTCACTGCTGTCCGAAAGTGGNAATCATGCGCCAGGCGCGGGGT 900
OY 357 rPThrArgAspGlyLysTyrAla-TrrPalMetPheLeu---AspArgProGlnGlnTrp 375
Db 901 GGACCCGGGATGGCAAAATACGCTCGNCATGTGCTCGGACCGGCGCCACAGCAGTGG 960
OY 376 LeuGlnLeu---ValLeuLeuProPro 383
Db 961 GCTCCAGNCTGTTCTCTTCCCTCC 987

RESULT 4
BO877413
LOCUS
DEFINITION
ABENCOURT_8049748 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5083674
5', mRNA sequence.
BO877413
VERSION
BO877413.1 GI:22269421
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 921)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL COMMENT

Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LICM2314 row: d column: 11
 High quality sequence stop: 651.

FEATURES

source
 1. 921
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."

BASE COUNT 205 a 266 c 266 g 184 t

ORIGIN

Alignment Scores:

Pred. No.: 6,7e-171 Length: 921
 Score: 1516.00 Matches: 294
 Percent Similarity: 97.37% Conservative: 2
 Best Local Similarity: 96.71% Mismatches: 4
 Query Match: 32.63% Indels: 4
 DB: 14 Gaps: 0

US-09-976-674-3 (1-863) x BQ877413 (1-921)

QY 199 AAlapheserPhellleasnanserAspleutrPyAlalaanllegluThrglyGlu 218
 Db 1 GCCTCTCTCTCTTCATCAT -AACAGCGACCTGTGGTGGCCAAATCGACAGCGGCG 59
 QY 219 GAlaargargleutPhrPhecySHISGLnglyUeuserasnValleuAapaspProlySer 238
 Db 60 GAGCGGCGGCTGACCTTCGCACCAAGTTATCCAAATGCTCGATGACCCCAAGTCT 119
 QY 239 AAlaglyValAlaThrPhenaIlleglIngluPhaSPAtrgPhetrGlyTrrPTrp 258
 Db 120 GCGGCTGTGGCACTTCGTCATACAGAGAGTTGACCGCTTCACTGGTATGGTGG 179
 QY 259 CysProThrAlaSerTrpGluGlySerGluGlyLeuLysThrLeuAqgIlleUeUrglu 278
 Db 180 TGGCCCCACACCTCTCTGGAGAGTTTCAGAGGCGCTCAAGACGCTGCAATCTGTATGAG 239
 QY 279 GluValaSpGluSerGluValGluValIlleHISValProserProAlaLeuGluAarg 298
 Db 240 GAAGTCATAGTCGAGGTGGAGGTCAATCACTCCCTCTCTGCGTCAGAGAAAGG 299
 QY 299 LysThrAspSerTrgArGlyTrProArGThrglySerLysAsnProLysIlleAlaLeuLys 318
 Db 300 AAGCGAGACGTCATTCGTCGTCAGACAGGACGCAAGAAATCCCAAGATTGCTTGA 359
 QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysIlleValSerThrGlnGluLysGluLeu 338
 Db 360 CTGGCTGAGTTCCAGACTGCAGCGAGGCAAGATTCCTCTCGACCCAGAGAAAGACTG 419
 QY 339 ValGlnProPheSerSerLeuPheProLysValGluTrIleAlaAargAlaGlyTrrPThr 358
 Db 420 GTGACGCCCTTCACGCTGCTGTCGCCAAGGTGAGTACATCGCCAGAGCGGGGTGAGCC 479

QY 359 ArgAspGlyLysTrgArAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeu 378
 Db 480 CGGATGGAATAATACGCGCTGGCCATGCTCTGGACCGGCCAGCAGTGGCTCCAGCTTC 539
 QY 379 ValLeuLeuProProAlaLeuPheIlleProSerThrGluAsnGluGlnGlnAargLeuAla 398
 Db 540 GTCTCTCTCCCGCGCGCTGTTATCCAGGACACAGAAATGAGGAGCGGCTGAGCC 599
 QY 399 SerAlaArgAlaValProArgAsnValGlnProTrgValValTrgGluValThrAsn 418
 Db 600 TCTGCCAGAGCTGTCCCGCAGAAATGTCACAGCCGTCATGCTGTCACGAGGTACCAAC 659
 QY 419 ValTrpLeuAsnValHisAspIlePheTrpProPheProGlnserGluGluAspGlu 438
 Db 660 GTCTGATCATATGTCATGACATCTTCATCCCTTCCCAATCAGAGGAGAGAGAGAG 719
 QY 439 LeucylsPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTrpLysValThr 458
 Db 720 CTCTGCTTCTCCCGCGCAATGAATGCAAGCGGCTTGTGCCATTTGTACAAAGTCACC 779
 QY 459 AlValLeuLysSerGlnGlyTrgAspTrpSerGluProPheSer-ProGluGluAspGlu 478
 Db 780 GCGCTTTTAAATATCCAGGCTACGATGAGTAGACCTTCACGCCCGGGGAAATGGA 839
 QY 478 uphelysCysProIleLysGluGluLe-AlaLeuThrSerGlyGlu-TrpGluValLeu 497
 Db 840 ATTTAACTGCCCATTTAAGGAAGATTTGGCTGTGACACAGGTCGAAATGGAAAGCTTT 899
 QY 498 AlAarg 499
 Db 900 GGCCGA 905

RESULT 5
 BQ684956
 LOCUS
 DEFINITION
 5', mRNA sequence.
 ACCESSION
 BQ684956
 VERSION
 BQ684956.1 GI:21810272
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 932)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS
 National Institutes of Health, Mammalian Gene Collection (MGC).
 TITLE
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LICM2393 row: g column: 16
 High quality sequence stop: 710.

FEATURES

source

1. 932
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6250407"
 /clone_id="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 209 a 265 c 269 g 188 t 1 others

Alignment Scores:

Pred. No.: 1,81e-170 Length: 932
Score: 1512.50 Matches: 295
Percent Similarity: 96.45% Conservative: 4
Best Local Similarity: 95.16% Mismatches: 9
Query Match: 32.55% Indels: 4
DB: 14 Gaps: 1

US-09-976-674-3 (1-863) x B0684956 (1-932)

QY 199 AlapheserPhelieasnsSerspieutryValaalaanileguthGlyclu 218
DB 1 GCGTTCTTCTCTCATCAATACAGCGACCTGTGGCCAAATCGAAGACGCGAG 60
QY 219 GluArgArgLeuthrPhecySHISGLnglyLeuserasnaValleuaspaaprolysSer 238
DB 61 GAGCGGGGCTGACCTTCTGCGACCAAGTTTATCCAAATGCTGGATGACCCCAAGTCT 120
QY 239 AlaglyValAlathrPhevalilegIngluUpheasparqhetnrglyTyrTPTPr 258
DB 121 GCGGGTGTGGCCACTTCGTCATACAGGAAGTTCCAGCCCTCAGTGGTACTGTGG 180
QY 259 CysProthralaserTpgluglySerluglyLeulysthrleuarqgleLeutyGlu 278
DB 181 TGCCCCACAGCTCTCTGGAGGTTTCAGAGGCTCAAGACGCTGCAATCTGTATGAG 240
QY 279 GluValaspgluserGluValilehistaProserProAlaleugluuarg 298
DB 241 GAATCGATGAGTCCGAGGTGAGGTGATTCACCTCCCTCTGCGCTAAGAAAGG 300
QY 299 LysThrAsperTyrArgTyrProArgThrglySerIysAsnProlyslleAlaLeulys 318
DB 301 AAGACGAGCTCGATCGGTACCCAGGACGACGACGAAGATCCCAAGATTGCTTGA 360
QY 319 LeuAlaelphegInthrAspserGlnGlyllyleValSerThrgIngluLysGluLeu 338
DB 361 CTGGCTAGTTCACAGCTGACAGCGGCGCAAGATCTCTGACCCAGGAGAAAGAGCTG 420
QY 339 ValGlnProPheSerleuPheProLysValGlnTyrIleAlaArgAlaGlyTyrThr 358
DB 421 GTGACGCTCTCAGCTGCTGTCGAGAGGTGAGTACATGCGCAGGCGGCTGAGCC 480
QY 359 ArgAspIyLysTyrAlaTPrAlaMetPheLeuAsparqProGlnGlnTPrLeuGlnLeu 378
DB 481 CGGATGTGCAAAATACGCTGGCGCATGTCTCGAACCGGCCACAGTGGCTCCAGCTC 540
QY 379 ValLeuLeuProProAlaLeuPheleProSerThrgluAsngluGluAlaLeuAla 398
DB 541 GTCTCTCTCCCGCGCGCTGTTCATCCGACAGAGATGAGAGACAGCGCTAGCC 600
QY 399 SerIlaArgAlaValProAlaGAsnValGlnProTyrValValTyrGluGluValThrAsn 418
DB 601 TCTGCCAGAGCTGTCGCCAGGAATGTCAGCGGATGTGTGTCGAGGAGAGTACCAAC 660
QY 419 ValTPrIleAsnValIleAspIlePheTyrProPheProGlnSerGlu-GlyGluAspG1 438
DB 661 GTCTGATCAATGTTATGACATCTTCTATCCCTCCCAATCAGAAAGAGAGAGAGA 720
QY 438 uLeuCyPheLeuArgAlaAsngluCysLysThrglyPheCysHisLeutyTyrValTrh 458
DB 721 GCTCTGCTTCTCCGCCCAATGATGATCAAGACGCGCTTGTGCATTTGTCAAAAGTCAC 780
QY 458 rAlaValLeuLysSerGlnGlyTyrAspTPrSerGluProPheSerProGluGluAspG1 478
DB 781 CGCCGTTTAAAT-TCCAAGGGCTACCAATGGAGTGA-CCCTTGGCGCCCGGGGAGAGAGA 838
QY 478 uPheLysCysProIleLysGluIleAlaLeuThrSerGlyGluTPr--GluValle 497
DB 123 GTGTCAGCTATGAGGGCGCGGAGATGCTAGCGCTCACACAGCGCGGCTTCCCAT 182

DB 839 ATTAAGTCCCAATTAAGGAGAGATGCTGTGACCCAGCGGAAATGGGAGAGTTT 898
QY 497 uAlaArgHisGlySerIyIleTPrVal 506
DB 899 GCGGAGCAGCGCTCAAAATTTGGGTT 926

RESULT 6
LOCUS B0678015
DEFINITION B0678015 982 bp mRNA linear EST 15-JUL-2002
ACCESSION AGENCOURT_8034317 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213986
VERSION B0678015
KEYWORDS B0678015.1 GI:21790694
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
TITLE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS 1 (bases 1 to 982)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LICM2380 row: 3 column: 03
High quality sequence stop: 683.

FEATURES
source
1..982
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6213986"
/clone_id="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site: 1; XhoI; Site: 2;
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(6). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 202 a 310 c 290 g 179 t 1 others

Alignment Scores:

Pred. No.: 2.69e-167 Length: 982
Score: 1486.50 Matches: 295
Percent Similarity: 93.40% Conservative: 2
Best Local Similarity: 92.77% Mismatches: 17
Query Match: 32.00% Indels: 6
DB: 14 Gaps: 1

US-09-976-674-3 (1-863) x B0678015 (1-982)

QY 489 LeuthSerGlyGluTPrGluValleuAlaArgHisGlySerLysIleTPrValaAsnglu 508
DB 3 CTGACCAAGCGGTGATGGAGGTTTGGCGAGCAGCGCTCAAGATCTGCGTCAATGAG 62
QY 509 GluThrLysLeuValTyrPheGlnGlyThrLysAspTPrProLysGlnHisLeuty 528
DB 63 GAGACCAAGCGGTGCTCTCCAGGACCAAGAGACGCGCGTGGAGACACACTTAC 122
QY 529 ValValSerTyrGluAlaAlaGlyGluLeuValArgLeuThrTPrProGlyPheSerHis 548
DB 123 GTGTCAGCTATGAGGGCGCGGAGATGCTAGCGCTCACACAGCGCGGCTTCCCAT 182

QY 549 SerCysSerMetSerGlnAsnPheAspMetPheValSerHisTyrSerSerValSerThr 568
 |||||
 Db 183 AGCTGCTCATGAGCCAGAACTTGACATGTCGTACAGCCACTACAGAGGTGAGCAGC 242
 QY 569 ProProCysValHisValTyrTyrLeuSerGlyProAsnSpasPspProLeuHisLysGln 588
 |||||
 Db 243 CCGGCGCTGCTGACGCTACAGACTGAGCGGCCCCACAGACGCCCTTCACAAACAG 302
 QY 589 ProArgPheTrpPalaSerMetMetGluAlaAlaSerCysPspProAspTyrValProPro 608
 |||||
 Db 303 CCGCGCTTCTGGGCTACAGATGATGAGGACAGCCAGCTGCCCCCGGATTTATGTTCTTCCA 362
 QY 609 GluIlePheHisPheHisTrpArgSerAspValArgLeuTyrGlyMetIleTyrLysPro 628
 |||||
 Db 363 GAGATCTTCATTTCCACAGCGCGCTGGATGTGGCGCTACAGGCATGATCTACAAAGCCC 422
 QY 629 HisAlaLeuGlnProGlyLysLysHisProThrValLeuPheValTyrGlyGlyProGln 648
 |||||
 Db 423 CAGGCTTGGACGCCAGGGAAGACCCACCGCTCTTGTATGTGAGGCGCCAG 482
 QY 649 ValGlnLeuValAsnAsnSerPheLysGlyIleLysTyrLeuArgLeuAsnThrLeuAla 668
 |||||
 Db 483 GTGACGCTGGTGAATTAATCTCTTCAAGGCAATCAAGTACTTGGCGCTCAACACACTGGCC 542
 QY 669 SerLeuGlyTyrAlaValAlaValIleAspGlyArgGlySerCysGlnArgGlyLeuArg 688
 |||||
 Db 543 TCCCTGGGCTACGCGCGTGTGATTTAGAGCGAGCGGCTCTGTACGCGAGGCTTCGG 602
 QY 689 PheGluGlyAlaLeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGly 708
 |||||
 Db 603 TTGGAAAGGGGCGCTGAAACCAATATGGCCAGGTGAGATCGAGGACCGAGGTGAGGCGC 662
 QY 709 LeuGlnPheValAlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGly 728
 |||||
 Db 663 CTGACGTTCTGGCGCGAAGTATGGCTTATGCACTGAGCCAGTGGCATTCATGAGC 722
 QY 729 TrpSerTyrGlyGlyPheLeuSerLeuMetGlyLeuIleHisLysProGlnValPheLys 748
 |||||
 Db 723 TGGGCTACGGGGGCTCTGCTGCTCATGGGCTAATGCC -AAGCCCCAGNTGTTCAAG 781
 QY 749 ValAlaIleLeuGlyAla -ProValThrValTrpMetAlaTyrAspThrGlyTyrThrGlu 768
 |||||
 Db 782 GGGGCGCTACGGGGTGGCGCCGCTGACCGTGTGATGGCTTACACACAGGATCACTGA 841
 QY 768 uArg-TrpMetAspValProGluAsnAsnGlnHisGlyTyrGluAlaGlySerValAlaIle 788
 |||||
 Db 842 GCGGCTACATGAGAACGTCCTGAGAACACAG -CAGGCGCTTAGAAGGGGGGCTTCGGGGC 900
 QY 788 euHisVal---GluLysLeu-ProAsnGluProAsnArgLeuLeuIle 802
 |||||
 Db 901 CCTGACGCGGGGAAACCTGCCCCAGAACCCACCCTGTTGCTT 948
 RESULT 7
 BQ897707 880 bp mRNA linear EST 16-AUG-2002
 LOCUS BQ897707
 DEFINITION AGENCYOUTR_8061873 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208407
 5', mRNA sequence.
 ACCESSION BQ897707
 VERSION BQ897707.1 GI:22289721
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 880)
 AUTHORS NIH-MGC <http://mgi.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LILCM2366 Row: a Column: 16
 High quality sequence stop: 704.
 Location/Qualifiers
 1. 880
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6208407"
 /clone_lib="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site1: XhoI;
 Site2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAAGAG(6). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH-MGC library."

BASE COUNT 194 a 260 c 249 g 176 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,69e-165 Length: 880
 Score: 1468.00 Matches: 287
 Percent Similarity: 97.95% Conservative: 0
 Best Local Similarity: 97.95% Mismatches: 0
 Query Match: 31.60% Indels: 6
 DB: 14 Gaps: 0

US-09-976-674-3 (1-863) x BQ897707 (1-880)
 QY 199 AlaphPheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluThrGlyGlu 218
 Db 1 GCCTTCTTCTCTTCATCATTAACAGCGACCTGTGGTGGCAACATCAGACAGCGCAG 60
 QY 219 GluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSer 238
 |||||
 Db 61 GAGCGGGGCTGACCTTCTTCCACCAAGTTTATCCATGTCTCGGATGACCCCAAGTCT 120
 QY 239 AlaGlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrp 258
 |||||
 Db 121 GCGGCTGTGGCCACCTTCGTCATACAGAGAGATGTCAGCGCTTCACTGGGTACTGTTGG 180
 QY 259 CysPProThrAlaSerTrpGluGlySerGlyGlyLeuLysThrLeuArgIleLeuTyrGlu 278
 |||||
 Db 181 TGCCCCACAGCCCTCTGGGAAGTTTCAGAGGGCTCAAGACGCTGGCAATCCGTATGAG 240
 QY 279 GluValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArg 298
 |||||
 Db 241 GAGTGCATAGTCCGAGGGAGGTGAGTATTCAGCTCCCTCTCTGCGTAGAAGAAAG 300
 QY 299 LysThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLys 318
 |||||
 Db 301 AAGACGACCTCGATGCGTACCCAGGACAGGACGACAGAAATCCCAAGATTGCTTGAAA 360
 QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeu 338
 |||||
 Db 361 CTGCGTGAAGTTCCAGACTGCACACCGAGGCGCAAGATGCTTCGACCCAGAGAGAGCTG 420
 QY 339 ValGlnProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThr 358
 |||||
 Db 421 GTGACGCTTTCAGCTGCTGCTCCGAAAGTGAAGATCCCAAGGCGGGGTGAGACC 480
 QY 359 ArgAspGlyLysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpGluGlnLeu 378
 |||||
 Db 481 CCGGATGCAAAATACGCTGCGCATCTTCTGAGCGGCGCCACGACAGTGGCTCCAGCTC 540
 QY 379 ValLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGluArgLeuAla 398

```

|||||
Db 541 GTCTCTCCCGCCCGCTTTCATCCCGACAGAGATGAGGAGCAGCGCTAGCC 600
Qy 399 SerIaArGaIaValProAIdgAsnValGInProTyYrValValTYrGluGluValThrAsn 418
Db 601 TCTCCAGAGGTGTCGCCAGGAATGTCCAGCCGTATGTGTAGCGAGAGTCAACCAAC 660
Qy 419 ValTrpIleAsnValHisAspIlePheTyPProPheProGInserGluGluAspGlu 438
Db 661 GTCTGGATCAATGTTCAATGACATCTTCTATCCCTTCCCATCAAGAGAGAGAGAG 720
Qy 439 LeuCySPheLeuAaGAlaAsnGluCysLysThrGlyPheCysHisLeuTYrLysValThr 458
Db 721 CTCTGCTTTCCTCCGCGCAATGCAAGACCGGCTTCTGCCAATTGTACAAAGTCAAC 780
Qy 459 Ala-ValLeuLysSerGln-GlyTYr-AspTrpSerGlu-ProPheSer-ProGlyGluA 477
Db 781 GCCCGTTTAAATCCAGAGGGCTACCGATGTGAGTGAAGCCCTTCAGCCCGGGGAG 840
Qy 477 spGluPheLysCys-ProIleLysGlu 485
Db 841 ATGAATTTAAGTCCCGCCCATTAAGGAA 867

RESULT 8
BO895618 943 bp mRNA linear EST 16-AUG-2002
LOCUS ABENOCURT_8753137 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6394542
DEFINITION 5', mRNA sequence.
ACCESSION BO895618
VERSION BO895618.1 GI:22287632
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 943)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@pilot.nih.gov
Tissue Procurement: Mark Macdonald, Ph.D. and Nancy L. Freeman,
Ph.D.
CDNA Library Preparation: Resgen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13887 row: m column: 07
High quality sequence stop: 729.

FEATURES
source
Location/Qualifiers
1..943
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6394542"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olocysts; Vector: PCMV-SPORT6.1.cdb;
Site:1: ECoRV; Site:2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by Resgen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 216 a 276 c 256 g 193 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 7 48e-162 Length: 943
Score: 1441.00 Matches: 260
Percent Similarity: 93.07% Conservativeness: 22
Best Local Similarity: 85.81% Mismatches: 20
Query Match: 31.02% Indels: 1
DB: 14 Gaps: 0

```

```

US-09-976-674-3 (1-863) x BO895618 (1-943)
Qy 400 AlaArGaIaValProAIdgAsnValGInProTyYrValValTYrGluGluValThrAsnVal 419
Db 20 GCCAGAGCCGTCGCCAAGGAATGTCCAGCCCTTTCATCTATGAAGAAGTCAACCAATGTC 79
Qy 420 TrpIleAsnValHisAspIlePheTyPProPheProGInserGluGluAspGlu 439
Db 80 TGGATCAACGTCACGACATCTTCACCGCTTCTCAAGGCGCCAGAGAGCTTT 139
Qy 440 CysPheLeuAaGAlaAsnGluCysLysThrGlyPheCysHisLeuTYrLysValThrAla 459
Db 140 TGTTCCTCTGTCGCCAAGAGATGCAAGAGTGGCTTCTCCACCTGACAGGCTCAGAGT 199
Qy 460 ValLeuLysSerGlnGlyTYrAspTrpSerGluProPheSerProGlyGluAspGluPhe 479
Db 200 GAACTTAAACCAAGAGACTATGACTGACGAGAACCCCTACAGAGATGAGTTT 259
Qy 480 LysCysProIleLysGluGluIleAlaLeuThrSerGlyGluTrpGluValLeuAlaArg 499
Db 260 AAGTGCCTCCATCAAGAGAGAGGTGCGCTGACAGTGGCGAGTGGAGGTCTTGCGAGG 319
Qy 500 HisGlySerLysIleTrpValaAsnGluGluThrLysLeuValTYrPheGlnGlyThrLys 519
Db 320 CATGGCTCCAGATCTGGGTCAACGACGACGACGAAAGCTGGTGTACTTCAAGGTACAAAG 379
Qy 520 AspThrProLeuGlnHisLeuTYrValValSerTYrGluAlaAlaGlyGluVal 539
Db 380 GACACACCGCTGGAACATCACCTGTATGTGTGACGTGACGAGTCAAGAGGAGTCTGTG 439
Qy 540 ArgLeuThrTrpProGlyPheSerHisSerCysSerMetSerGlnAsnPhaSerPhe 559
Db 440 CGGCTCACACGCTCGGCTTCTCCACAGCTGTCCACAGACCCAGAGCTTGACATGTC 499
Qy 560 ValSerHisTYrSerSerValSerThrProProCysValHisValTYrLysLeuSerGly 579
Db 500 GTGAGTCACTACAGAGAGTGTAGCAGCAGCCCTGTGTACATGTGTACAACTAGAGCGG 559
Qy 580 ProAspAspAspProLeuHisLeuGlnProAlaArgPheTrpAlaSerMetGluAla 599
Db 560 CCCGATGTAGACCCACCTGACAGACCAACCCGCTTGGCGCAGATGATGAGGACAC 619
Qy 600 SerCysProProAspTYrValProProGluIlePheHisPheHisThrArgSerAspAla 619
Db 620 AATGGCCCCAGACTATGTGCCCCCTGAGATCTTCCACACCCGTCGACAGAGCTG 679
Qy 620 ArgLeuTYrGlyMetIleTYrLysProHisAlaLeuGlnProGlyLysHisProThr 639
Db 680 CAGCTCTACGGCATGATGTACAGGACACACCCCTGCACAACTGAGAGAGACACCCCACT 739
Qy 640 ValLeuPheValTYrGlyGlyProGlnValGlnLeuValaAsnSerPheLysGlyIle 659
Db 740 GTGCTCTTGTCTATGGGGGCCCAAGGTGAGTGGGAAACACTCTTTTAAAGCATC 799
Qy 660 LysTYrLeuAaGlyLeuAsnThrLeuAlaSerLeuGlyTYrAlaValVal-ValIleAspGlu 679
Db 800 AAATACCTGGCGGCTAAATACACTGTCATCTGGCTGTGGTGTGCTGAGTGCAGAGG 859
Qy 679 YARGLYSerCysGlnAaGlyLeuArgPheGluGluValaLeuLysAsnGlnMetGlyGlu 699
Db 860 TCGGGGCTCTGTGTACGGGGGCTGCACTTGTGAGGGGGGCGCTGAAATCAATAGGGGCC 919

RESULT 9
BO889360 879 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOCURT_8061878 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208551
DEFINITION 5', mRNA sequence.
ACCESSION BO889360

```

VERSION BQ889360.1 GI:22281374
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2366 row: g column: 16
High quality sequence stop: 709.
Location/Qualifiers
1..879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:6208551"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 194 a 256 c 252 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 1.14e-161 Length: 879
Score: 1439.00 Matches: 279
Percent Similarity: 96.58% Conservative: 3
Best Local Similarity: 95.55% Mismatches: 5
Query Match: 30.97% Indels: 5
DB: 14 Gaps: 2
US-09-976-674-3 (1-863) x BQ889360 (1-879)
QY 199 AAlaphePheSerPheIleAsnAsuSerAspLeuTrpValAlaAsnIleGluThrGlyGlu 218
|||||
Db 1 GCCTCTTCCTTCATCAATTAACAGCGACTGTGGGTGCCAATCAACGAGACGGGAG 60
QY 219 GluAtrGArgLeuThrPheCySHsGlnGlyLeuSerAsnValLeuAspAspProLysSer 238
|||||
Db 61 GACGGGGGCTGACCTCTCCACCAAGGTTATCCAAATGCTCCGATGACCCCAAGTCT 120
QY 239 AlAGlyValAlaThrPheValIleGlnGluGluPheAspArgPheThrGlyTyrTrp 258
|||||
Db 121 GCGGGGTGTGGCCACTTGCATACAGGAAGAGTTCACCGCTTCACTGGGTACTGGTGG 180
QY 259 CysProThrAlaSerTrpGluGlySerGlnGlyLeuLysThrLeuArgIleLeuTyrGlu 278
|||||
Db 181 TGGCCCCACACCTCTCTGGGAAGTTCAGAGGCTCAAGACCTGCGCAATCTCTATGAG 240
QY 279 GluValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArg 298
|||||
Db 241 GAAGTCATAGATCCGAGGTGAGAGTCATTCACGTCCTCCCTCGGTGAGAAAG 300
QY 299 LysThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLys 318
|||||
Db 301 AAGACGACTCGTATCGTACCCAGAGACAGCAAGAAATCCCAAGATTGGCTTGAAA 360

QY 319 LeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeu 338
|||||
Db 361 CTGGCTAGTTCACAGACTGACAGCGAGCAAGATCTCTCGACCCAGAGAAAGAGACTG 420
QY 339 ValGlnProPheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTyrPThr 358
|||||
Db 421 GTCCAGCCCTTACAGCTCGTGTCCCAAGAGTGAGTACATCCAGAGCGCGGTGGACC 480
QY 359 ArgAspGlyLysTyrAlaTProAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeu 378
|||||
Db 481 CGGAGTGGCAATACGCTGCGATCTCTGAGCCGCGCCAGCAGTGGTCCAGCTC 540
QY 379 ValLeuLeuProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAla 398
|||||
Db 541 GTCTCTCTCCCGCGCGCGCTTCATCCGACAGACAGAAATGAGAGAGCGGCTAGCC 600
QY 399 SerAlaArgAlaValProArgAsnValGlnProTyrValValTyrGluValThrAsn 418
|||||
Db 601 TCTGCCAGACTCTGCCAGGAATGTCACCGCATGTGTGTACGAGAGGTCCACCAC 660
QY 419 ValTrpIleAsnValHisAspIlePheTyrProPheProGlnSerGluGluAspGlu 438
|||||
Db 661 GTCTGGATCATGTTCATGACATCTTATCCCTTCCCAATCAGAGGAGAGAGAG 720
QY 439 LeuCysPheLeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThr 458
|||||
Db 721 CTCTGCTTCTCCGCGCGCATGATGACACCGGCTTCGCCATTGTGTAAGAAGTCAAC 780
QY 459 Ala-ValLeuLysSerGln-GlyTyrAsp-TrpSerGluProPheSer---ProGlyGlu 476
|||||
Db 781 GCCCTTTTAAATCCACGAGGCTACGATTGGATGATGAGCCCTTACGCCCGCGGGAG 840
QY 477 AspGluPhe---LysCysProIleLysGlu 485
|||||
Db 841 GATGAATTTTAAGTGGCCCATTAAGAA 870
RESULT 10
BM461814
LOCUS BM461814
DEFINITION 5', mRNA sequence.
ACCESSION BM461814
VERSION BM461814.1 GI:18510854
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12219 row: g column: 20
High quality sequence stop: 627.
Location/Qualifiers
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:5533867"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI;

BASE COUNT	188 a	269 c	245 g	164 t	Site 2: Saili: Cloned unidirectionally. Primer: Oligo dT
ORIGIN					Average insert size 2.1 kb.

BASE COUNT	188 a	269 c	245 g	164 t
ORIGIN				

Alignment Scores:

Pred. No.:	7,12e-161	Length:	866
Score:	1432.00	Matches:	276
Percent Similarity:	97.87%	Conservative:	0
Best Local Similarity:	97.87%	Mismatches:	2
Query Match:	30.82%	Indels:	4
DB:	13	Gaps:	0

US-09-976-674-3 (1-863) x BM461814 (1-866)

QY	40	GlySerArgLysTrpSerGlyLeuLeuValAsnLysAlaProHisAspPheGlnPheVal	59
Db	11	GGCAGCCGCAAGTACTCGGGCTCATTTGTCACAAAGGGGCCCCACGACTTCCAGTTTGTG	70
QY	60	GlnLysThrAspLysSerGlyProHisSerHisArgLeuTrpTyrLeuGlyMetProTyr	79
Db	71	CAGAAAGCGGATGAGTCTGGGCCCATCTCCACCGGCTCTACTACTCGGGAAATGCCATAT	133
QY	80	GlySerArgGlnAsnSerLeuLeuTyrSerGluLeuPolysLysValArgLysGluAla	99
Db	131	GGCAGCCGAGAAAGTACCTCTCTACTAGATATCCCAAGAAAGCTCCGAAAGAGGCT	19
QY	100	LeuLeuLeuLeuSerTrpLysGlnMetLeuAspHisPheGlnAlaTrpProHisHisGly	111
Db	191	CTGTGTGCTCTGCTCCGGAAAGATGCTGGATCATTTTCCAGGGCCACGGCCCACTCATG	25
QY	120	ValTyrSerArgGlnGluLeuLeuLeuArgGluArgLysArgLeuGlyValPheGlyIle	133
Db	251	GTCACCTCTCGGGAGGAGGAGCTGCTGAGGGAGCGGAACCCCTGGGGGTCTTCCGCATC	31
QY	140	ThrSerTyrAspPheHisSerGlyLeuPheLeuPheGlnAlaSerAsnSerLeu	15
Db	311	ACCTCTACACTTCCACAGCAGAGTGGCTTCTTCCCTCCACAGGCCACACAGCTTC	37
QY	160	PheHisCysArgAspGlyLysAsnGlyPheMetValSerProMetLysProLeuGlu	17
Db	371	TTCCACTGCGCGCACGGCGGCAAGAGCGCTTCAATGGTGTCCCTTATGAACCGCTGGAA	43
QY	180	IleLysThrGlnCysSerGlyProArgMetAspProLysIleCysProHisAspProAla	19
Db	431	ATCAAGACCAAGTGCCTCAGGGCCCCGGATGGACCCCAAAATCTGCTTCCGCACCTGCC	49
QY	200	PhePheSerPheIleAsnAsnSerAspLeuTrpValAlaAsnIleGluTrpGlyGlu	21
Db	491	TTCTCTCTCTCATCATATACAGCACTGGTGGGGGCCAACATCGAGACGCGAGGAG	55
QY	220	ArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeuAspAspProLysSerAla	23
Db	551	CGGGGGGTGACTTTCGCCACCAAGGTTTATTCATATGCTCGGATGACCCCAAGTCTGGG	61
QY	240	GlyValAlaIleThrPheValIleGlnGluLupPheAspArgPheThrGlyTrpTrpCys	25
Db	611	GGTGTGGCCACTTCTCTACTACAGAAAGTTCCAGCCCTTCACTGGGTACTGTGTGTC	67
QY	260	ProThrAlaSerTrpGlnLysSerGlyLeuLysTrpThrLeuArgIleLeuTyrGlnGlu	27
Db	671	CCCAAGAGCTCTCGGGAAGGTCAAGAGGGCTCAAGACGTCGCAATCTGTTAGAGAA	73
QY	280	ValAspGlnSerGlnValGlnValIleHisValProSerProAlaLeuGluLupArgL	29
Db	731	GTCGATGAGTCCCGAGAGTGGAGTATTTCACGTCTCCCTCTCGCTTAAAGAAAAAG	79
QY	299	LysThrAspSerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeu	317
Db	791	AGAGGAGCTGTATGCTGTACCCCAAGAACAGGCAAGATGCCCAAGATGGCTT	848
RESULT	11		
	90689588		

LOCUS	BO689588	872 bp	mRNA	linear	EST 15-JUN-2002
DEFINITION	AGENCOURT_8345928 NIH_MGC_110 homo sapiens cDNA clone IMAGE:6250957				
ACCESSION	BO689588				
VERSION	BO689588.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 872)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@b6-remail.nih.gov Tissue Procurement: ATCC				
FEATURES	cDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: http://image.lnl.gov Plate: LCM2394 row: n column: 14 High quality sequence stop: 654.				
SOURCE	Location/Qualifiers 1..872 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6250957" /clone_1id="NIH_MGC_110" /tissue_type="ductal carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: pancreas; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."				
BASE COUNT	192 a	258 c	247 g	174 t	1 others
ORIGIN					
Alignment Scores:					
Prod. No.:	3.48e-158	Length:	872		
Score:	1410.00	Matches:	274		
Percent Similarity:	96.15%	Conservative:	1		
Best Local Similarity:	95.80%	Mismatches:	6		
Query Match:	30.35%	Indels:	5		
DB:	14	Gaps:	2		
US-09-976-674-3 (1-863) x	BO689588 (1-872)				
OY	199	AlaPhePheSerPheIleAsnAsnSerAspLeuTPValAlaAsnIleGluThiGlyu	218		
Db	1	GCCTCTCTCTCTCATCAATACACGACGACCTGTGGTGGCCACATCGACACAGCGAG	60		
OY	219	GluArgArgLeuThrPheCysHisIscInglyLeuSerAsnValIleuAspAspProLysSer	238		
Db	61	GAGCGCGCGCGTACCTTGTCCACCCAAAGGTTATCAATATGCTCTGGATGACCCCAAGCT	120		
OY	239	AlaGlyValAlaThrPheValIleIleIngluIngluPheAspArgPheThrGlyTyrTyrTrp	258		
Db	121	GGGGGTGGGGCCACTTGTCTATACAGGAAGAGTTCGACCCGCTTACAGGGGACTGGGG	180		
OY	259	CysProThrAlaSerTrrGluGlySerGluInglyLeuLysThrIleuArgIleLeuTyrGlu	278		
Db	181	TGCCCCACAGGCTCTGGGAAGGTTCAAGAGGGCCCAAGAGCGTCCGATCTGTATGAG	240		
OY	279	GluValAspGluSerGluValGluValIleHisValProSerProAlaLeuGluGluArg	298		
Db	241	GAAGTCGATGAGTCCGAGAGTGGAGGTCAATACGCGCCCTCTCGCGCTGAAGAAGG	300		


```

Db 756 TGTGGGACCTTCATACAGAGAGAGTGCACCCGCTTCACTGGTACTGGTGGCC 815
Qy 260 rothAla-SerTrpGluGlySerGluGlyLeuYstHleu-AngIIleuTyrglu 279
Db 816 CCACAGCCCTCTGGGAGGTTCAAGAGGCGCTCAGACGCTGCCAATCTGTATGTAGA 875
Qy 279 uValaspluSer 283
Db 876 AGTCATGAGTCC 888

RESULT 13
LOCUS BM915935
DEFINITION AGNCOURT.6639540 NIH_MGC_41 Homo sapiens CDNA clone IMAGE:5482205
5' mRNA sequence.
ACCESSION BM915935
VERSION BM915935.1 GI:19366314
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1051)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2007 row: 0 column: 06
High quality sequence stop: 677.

FEATURES
source
location/Qualifiers
1..1051
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5482205"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Skin; Vector: pORB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(g). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

BASE COUNT 233 a 329 c 300 g 189 t
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-156 Length: 1051
Score: 1395.00 Matches: 280
Percent Similarity: 92.46% Conservative: 2
Best Local Similarity: 91.80% Mismatches: 12
Query Match: 30.03% Indels: 11
Gaps: 3
DB: 14

US-09-976-674-3 (1-863) x BM915935 (1-1051)
Qy 114 AlaThrProHisGlyValTyrSerArgGluGluLeuLeuArgGluArgLysArg 133
Db 2 GGCAGCCCCACCATGGGTTCTACTCTCGGAGAGAGAGACTGCTGAGGAGCGGAACCC 61
Qy 134 LeuGlyValPheGlyIleThrSerTyrAspPheHisSerGluSerGlyLeuPheLeuPhe 153

```

```

Db 62 CTGGGGGCTCTTCGGCATCACTCTACAGCTTCACAGGAGAGTGGCTTCTCTTC 121
Qy 154 GlnAlaSerAsnSerLeuPheHisCysArgAspGlyGlyAsnGlyPheMetAlaSer 173
Db 122 CAGGCGAGCAACAGCCTCTTCCACTGCGCGGAGGCGCAAGAGCGCTTCATGGTGTCC 181
Qy 174 PrometLysProLeuGluIleLeuYstHleuGlySerGlyProArgMetAspProLysIle 193
Db 182 CCTATGAACCCCTGGAAATCAAGACCCAGTCTCAGGCCCCGGATGGACCCCAAAATC 241
Qy 194 CysProAlaAspProAlaPhePheSerPheIleAsnAsnSerAspLeuTyrValAlaAsn 213
Db 242 TCCCTGCGCGAGCCTTCCCTTCTTCATCAATAAACAGGACCTGTGGGTGGCAAC 301
Qy 214 IlegIuThrGlyGluGluArgArgLeuThrPheCysHisGlnGlyLeuSerAsnValLeu 233
Db 302 ATCGAGACAGGGAGAGGCGGCTGACCTTCCACCAAGGTTTATCCAAATGCTCG 361
Qy 234 AspAspProLysSerAlaGlyValAlaThrPheValIlegIuGluGluPheAspArgPhe 253
Db 362 GATGACCCCAAGTCTCGGGGTGGCCACCTTCGTCATACAGAAAGATTCACCGCTTC 421
Qy 254 ThrGlyTyrTrpTrpCysProThrAlaSerTrpGlySerGluGlyLeuYstHleu 273
Db 422 ACTGGGTACTGGTGGTGGCCACAGCCTCTGGAAAGTTACAGAGGCGCTCAAGACGCTG 481
Qy 274 ArgIIleuTyrgluGluValAspGluSerGluValGluValIleHisAlaProSerPro 293
Db 482 CCAATCTGTATGAGGAAGTCAATGATGTCGAGGTGAGTCAATTCACCTCCCTCTCT 541
Qy 294 AlaLeuGluGluArgLysThrAspSerTyrArgTyrProArgThrGlySerLysAsnPro 313
Db 542 GCGCTAGAAGAAAGAGAGAGGAGACTGTATCGGTACCCAGAGACAGCAAGAAATCC 601
Qy 314 LysIleAlaLeuLysLeuAlaGluPheGlnThrAspSerGlnGlyLysIleValSerThr 333
Db 602 AAGATTGCTCTGAATGCTGCTGATTCAGACTGACAGCCAGGCAAGATCGTCTGACCC 661
Qy 334 GlnGluLysGluLeuValGlnProPheSerSerLeuPheProLysValGluIyrIleAla 353
Db 662 CAGGAGAAAGAGACTGTGTGAGCCCTTACCTGCTTCCAGAAAGTGGGT-----ACT 715
Qy 354 ArgAlaGlyTyr-----ThrArgAspGlyLysTyrAla-TyrAlaMetPheLeuAs 370
Db 716 CCGCAGAGGGGGCGGGGTGGAACCCCGAATGCGAAATAGCCCTGGCCATGCTCTGGA 775
Qy 370 parG-ProGlnGlnTrpLeuGlnLeuValLeuLeuProPro-AlaLeuPheIleProSer 389
Db 776 CCGGCCCCAGCAGATGGCTCCAGCTGCTCTCTCCCGGGGCTGTTCATCCGAGAC 835
Qy 390 ThrGlu--AsnGluGluGlnArgLeu---AlaSerAlaArgAlaValProArgAsnValG 408
Db 836 CCCAAGAATAGAGAGAGCGGCTTAACCTCTGCGCAAGTGGCCCAAGAAATGTCC 895
Qy 408 InPro 409
Db 896 AACCC 900

RESULT 14
LOCUS B1690774
DEFINITION 603314392F1 NCI_CGAP_Mam6 Mus musculus CDNA clone IMAGE:5354182 5',
mRNA sequence.
ACCESSION B1690774
VERSION B1690774.1 GI:15653403
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 821)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

```


JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11900 row: h column: 23
High quality sequence stop: 821.

FEATURES
SOURCE

BASE COUNT	183 a	237 c	234 g	167 t
ORIGIN				

Alignment Scores:	
pred. No.:	5,53e-155
Score:	1383.00
Percent Similarity:	97.81%
Best Local Similarity:	94.16%
Query Match:	29.77%
DB:	13
length:	821
Matches:	258
Conservative:	10
Mismatches:	4
Indels:	2
Gaps:	0

US-09-976-674-3 (1-863) x BI690774 (1-821)

QY	473	SePProGluGluAspGluPhePheCysProIleuGluGluIleAlaLeuThrSerGly	492
Db	4	AGCCCTACAGAAAGATGAATTAAATGCCCCCTCAAGAGAGAGTGGCCCTGACCAATGGC	63
QY	493	GLUTPRGluValLeuAlaAlaArgHisGlySerLysIleThrValAsnGluGluThrLysLeu	512
Db	64	GAGTGGGAGGCTCTGTGGAGGCATGGCTCCAAAGATCTGGGTCAAGCAGCAAGCAACCTG	133
QY	513	ValTYRPhenGlnGlyThrLysAspThrProIeuGluHisIleuTYRValValSerTYR	532
Db	124	GTTGACTTTCAAGGTACAAAGGACACACCGCTGGAATCATCCCTCTATGTGGTGAAGTAC	183
QY	533	GluAlaAlaGlyGluIleValArgLeuThrThrProGlyPheSerHisSerCysSerMet	552
Db	184	GAGTCACAGAGCGAGATCGTGGCGCTCACCCGCTGGCTTCTCCACAGCTGTCATG	243
QY	553	SerGlnAsnPheAspMetPheValSerHisIlyrTerSerValSerThrProProGlyVal	572
Db	244	AGCCAGAGCTTCGCATGATGTTGTAGTGCATACAGCAGTGGACACGCCACCCCTGTGA	303
QY	573	HisValTYRlysLeuSerGlyProAspAspArgProIeuHisIlyGlnProArgPheThr	592
Db	304	CATGTGTACAAAGTGGAGGGGCCCGAGATGATGCCACTGTGCACAAAGCAACCGCTTCTGG	363
QY	593	AlaSerMetMetGluAlaAlaSerCysProProAspTYRValProProGluIlePheHis	612
Db	364	GCCACGACATGAGGAGGAGCACCACAAATTTGCCCCCAACATATGTGGCCCCCTGGAGATCTTCAC	423
QY	613	PheHisThrArgSerAspValArgLeuTYRgluMetIleTYRlysProHisAlaLeuGln	632
Db	424	TTTCACACCCGTTGACAGAGCTGCACACTGTACGGCATGATGTTCAAAGCCACACACCTGTCAA	483
QY	633	ProGlyLysLysHisProThrValLeuPheValTYRgluGlyProGluValGlnLeuVal	652

Db	484	CCTGGGAGGAGACACCCACATGTGCTTTGTCTATTGGGGGGCCACAGGTGAGTTGGTG	543
Qy	653	AsnaSenSerPheLysGlyLeuLysTyrLeuArgLeuAsnThrLeuAlaSerLeuGlyTyr	672
Db	544	AACACACTCCTTTAAAGGGCATCAAAATACCTGGCGGCTAAATATACCTGGCATCCTTGGGGCTAT	603
Qy	673	AlaValValValIleAsnGlyArgGlySerGlyGlnArgGlyLeuArgPheGluGlyAla	692
Db	604	GCTGGGGGGGTATGATGATGATGGGGGCTCTCTCAGCGGGGCTCCTCACTGCAGGGGGCC	663
Qy	693	LeuLysAsnGlnMetGlyGlnValGluIleGluAspGlnValGluGlyLeuGlnPheVal	712
Db	664	CTGAAATAATCAATGGGGCAGGTGAGATTGAGCAGCAGGTGGAAGGCTGCAGTACGCG	723
Qy	713	AlaGluLysTyrGlyPheIleAspLeuSerArgValAlaIleHisGlyTyrPserTyrGly	732
Db	724	GCTGGGAAGTATGGCTTATTGACTTGGACCGAGTCGGCATTCATGGCTGGTCTTACGGC	783
Qy	733	GlyPheLeuSerLeuMet-GlyLeuIleHisLysPro	744
Db	784	CGGCTTCTCTTCATCATCATGAGGGGCTCATTCACCAAGCCA	821

RESULT	15
B0642814	
LOCUS	
DEFINITION	B0642814 902 bp mRNA linear EST 15-JUL-2002
ACCESSION	AGNCNCOURT_8285906 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:6292546
VERSION	5'', mRNA sequence.
KEYWORDS	B0642814 B0642814.1 GI:21766986
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
TITLE	Euxaiyotia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
COMMENT	1 (bases 1 to 902)
	NIH-MGC http://mgc.nci.nih.gov/ .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	unpublished (1999)
	Contact: Robert Strausberg, Ph.D.

Email: cgapds-remail.nlm.nih.gov
Tissue Procurement: Lou Straudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
plate: LNCM2494 row: c column: 11
High quality sequence stop: 637.

BASE COUNT	202	a	269	c	255	g	175	t	1	others
ORIGIN										
Alignment Scores:										
Pred. No.:	1.21e-153					Length: 902				
Score:	1372.50					Matches: 269				

Percent Similarity: 92.26% Conservative: 5
Best Local Similarity: 90.57% Mismatches: 11
Query Match: 29.54% Indels: 12
DB: 14 Gaps: 2

US-09-976-674-3 (1-863) x BQ642814 (1-902)

QY	282	GluSerGluValGluValIleHisValProSerProAlaLeuGluArgLysThrAsp	301
DB	3	GAGTCCGAGGTGAGGTCAATCAGTCCCTCTCCGCTAGAGAAAGAGACGGAC	62
QY	302	SerTyrArgTyrProArgThrGlySerLysAsnProLysIleAlaLeuLysLeuAlaGlu	321
DB	63	TCGTATCGGTACCCAGGACAGCAGCAAGATCCCAAGATTGCCTTGAAACTGGCTGAG	122
QY	322	PheGlnThrAspSerGlnGlyLysIleValSerThrGlnGluLysGluLeuValGlnPro	341
DB	123	TTCAGACTCACAGCCAGGCGCAGATCGTCTCGACCCAGGAGAGAGCTGGTGCACCCC	182
QY	342	PheSerSerLeuPheProLysValGluTyrIleAlaArgAlaGlyTrpThrArgAspGly	361
DB	183	TTCAGCTCGCTGTCCCGAAGGTGGAGTACATCGCCAGGCGCGGTGGACCCGGGATGGC	242
QY	362	LysTyrAlaTrpAlaMetPheLeuAspArgProGlnGlnTrpLeuGlnLeuValLeu	381
DB	243	AAATACGCCCTGGGCCATGTTCTCGACCGGCCCCAGCAGTGGCTCCAGCTCGTCTCCTC	302
QY	382	ProProAlaLeuPheIleProSerThrGluAsnGluGluGlnArgLeuAlaSerAlaArg	401
DB	303	CCCCCGGCCCTGTTCAATCCCGACACAGAAATGAGGAGCAGCGGCTAGGCTCTGCCAGA	362
QY	402	AlaValProArgAsnValGlnProTyrValValTyrGluValThrAsnValTrpIle	421
DB	363	GCTGTCCCCAGGAATCTCCAGCCGTATGTGTGTACGAGGAGGTCACCAAGCTCTGGATC	422
QY	422	AsnValHisAspIlePheTyrProPheProGlnSerGluGlyGluAspGluLeuCysPhe	441
DB	423	AATGTTTCATGACATCTTCTATCCCTTCCCCCAATCAGAGGAGAGGAGCTCTGCTTT	482
QY	442	LeuArgAlaAsnGluCysLysThrGlyPheCysHisLeuTyrLysValThrAlaValLeu	461
DB	483	CTCCGGGCCAATGAATCAGACAGCGGCTTCTGCCATTTGTACAAAGTCACCGCGGTTTA	542
QY	462	LysSerGlnGlyTyrAspTrpSerGluProPheSerProGlyGluAspGluPheLysCys	481
DB	543	AAATCCAGGGCTACGATTGGAGTGAGCCCTTCAGCCCCGGGGAGAGATGAATTAAGTGC	602
QY	482	ProIleLysGluGluIleAlaLeuThrSerGlyGluTrp-GluValLeuAlaArgHisG1	501
DB	603	CCCATTAAGGAAGAGATTGCTCTGACCAGCGGTGAATGGGAGGNTTTGGCGGACGCG	662
QY	501	YSerLysIleTrpValAsnGluGluThrLysLeuValTyrPheGlnGlyThrLysAspTh	521
DB	663	CTTCAAGATCTGGGTCAATGAAGACCAAGCTGGTGTACTTTTCAGGGCACCAGGACAC	722
QY	521	rProLeuGluHisHisLeuTyrValValSerTyrGluAlaAlaGlyGluIleValArgLe	541
DB	723	GGCGTGGAGCACCACTCTACGTGGTCACTATGAGCGCGCGGAGCAACCTAGGCCT	782
QY	541	uThrThr-ProGlyPheSerHisSerCysSerMet-SerGlnAsnPheAspMetPheVal	560
DB	783	CACCACCTCCGGGTCTTCCTTAAGCTGCTCCATGAAGCCAAAAATTC-----	830
QY	561	SerHisTyrSerSer-----ValSerThrProPro 570	
DB	831	CCACATGGTTCCGACGACACTTAGGGGATCTCTTAACACCCCCCCC 875	

Search completed: December 12, 2002, 13:53:23
Job time : 1966 secs

THIS PAGE BLANK (uspto)